

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:39:47 ; Search time 48 Seconds
(without alignments)
3025.613 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 2726

Sequence: 1 MSISSEVFNFLVRYLQESG.....GDKVGASASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2711	99.4	514	4 AAB95225	Aab95225 Human pro
2	2708	99.3	514	6 ABO07190	Abc07190 Human p53
3	2437	89.4	577	7 ADD14051	Add14051 Human src
4	2340	85.8	542	5 ABP41760	Abp41760 Human ova
5	2047	75.1	395	5 ABP51424	Abp51424 Human MDD
6	1907.5	70.0	584	4 ABG21351	Abg21351 Novel hum
7	1904.5	69.9	700	4 ABB60376	Abb60376 Drosophil
8	1137	41.7	208	4 AAO04385	Aao04385 Human pol
9	819	30.0	167	4 ABG21350	Abg21350 Novel hum
10	499.5	18.3	535	6 ABR52980	Abf52980 Protein s
11	374.5	13.7	414	6 ADA13321	Adal3321 Human int
12	370.5	13.6	481	4 ABB59486	Abb59486 Drosophil
13	370	13.6	479	3 AAY79678	Aay79678 Drosophil
14	368	13.5	411	4 ABB62260	Abb62260 Drosophil
15	367.5	13.5	485	4 AAB68284	Aab68284 Amino aci
16	366.5	13.4	485	4 AAB68516	Aab68516 Human GTP
17	366.5	13.4	485	4 AAB92844	Aab92844 Human pro
18	366.5	13.4	485	5 ABB97306	Abb97306 Novel hum
19	365.5	13.4	515	6 ABR53774	Abr53774 Protein s
20	364.5	13.4	514	2 AAR95881	Aar95881 WD-40 dom
21	359	13.2	352	3 AAG14893	Aag14893 Arabidops
22	359	13.2	352	3 AAG48119	Aag48119 Arabidops
23	359	13.2	352	7 ADB95040	Adb95040 A. thalia
24	359	13.2	484	4 AAB68282	Aab68282 Amino aci
25	356	13.1	349	3 AAG48120	Aag48120 Arabidops

26	356	13.1	349	3 AAG14894	Aag14894 Arabidops
27	348	12.8	339	2 AAW55857	Aaw55857 Human cia
28	348	12.8	339	2 AAW58888	Aaw58888 Human lip
29	348	12.8	339	7 ADE59693	Ades9693 Human pro
30	348	12.8	339	7 ADE59705	Ades9705 Human pro
31	348	12.8	339	7 ADE59697	Ades9697 Human pro
32	348	12.8	339	7 ADE59709	Ades9709 Human pro
33	348	12.8	339	7 ADE83378	Ades8378 Human pro
34	348	12.8	339	7 ADE89701	Ades9701 Human pro
35	348	12.8	339	7 ADE59713	Ades9713 Human pro
36	348	12.8	339	7 ADE57861	Ades7861 Human pro
37	348	12.8	450	6 ABB99407	Abb99407 Amino aci
38	348	12.8	451	6 ADA21145	Ada21145 Human sec
39	348	12.8	478	4 AAM93784	Aam93784 Human pol
40	348	12.8	478	6 ABB99402	Abb99402 Amino aci
41	348	12.8	521	4 ABB10141	Abb10141 Human cdn
42	348	12.8	521	5 ABF66728	Abf66728 Human pol
43	346.5	12.7	409	7 ADE54235	Ades4235 Human pro
44	344.5	12.6	409	2 AAR70002	Aar70002 OPDE 45 k
45	344.5	12.6	409	7 ADE58488	Ades8488 Rat Prote

ALIGNMENTS

RESULT 1
AAB95225

ID AAB95225 standard; protein; 514 AA.

XX AAB95225;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17352.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EF1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-003000253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 514 AA;
XX
XX
SO

OS Homo sapiens.
XX WO200299122-A1.
XX PPN
XX 12-DEC-2002.
XX PD
XX PD
XX 03-JUN-2002; 2002WO-US017382.
XX PF
XX 05-JUN-2001; 2001US-0296076P.
XX PR
XX 10-OCT-2001; 2001US-0328605P.
XX PR
XX 15-FEB-2002; 2002US-0357253P.
XX PR
XX (EXEL-) EXELIXIS INC.
XX PA
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX PPI
XX WPI; 2003-156859/15.
XX DR
XX N-PSDB; ACD13365.
XX DR
XX Identifying modulators of the p53 pathway for use in treating apoptotic
XX PPT or cell proliferation disorders, comprises screening for agents that
XX PPT modulate activity of a human ortholog of genes that modify the p53
XX PPT pathway in Drosophila.
XX PT
XX Example 2; Page 458-459; 678pp; English.
XX PS
XX The invention relates to identifying (M1) a candidate p53 pathway
XX CC modulating agent, by contacting an assay system comprising a purified HM
XX CC modulating agent, by contacting an assay system comprising a purified HM
XX CC polypeptide (human orthologue of genes that modify the p53 pathway in
XX CC Drosophila) or nucleic acid with a test agent under conditions, where but
XX CC for the presence of the test agent, the system provides a reference
XX CC activity, and detecting a test agent-biased activity of a cell comprising
XX CC Also included are modulating (M2) a p53 pathway of a cell comprising
XX CC contacting a cell defective in p53 function with a candidate modulator
XX CC that specifically binds to a HM polypeptide comprising an HM amino acid
XX CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
XX CC in a mammalian cell (comprising contacting the cell with an agent that
XX CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
XX CC a disease in a patient (comprising: (a) obtaining a biological sample
XX CC from the patient; (b) contacting the sample with a probe for HM
XX CC expression; (c) comparing the results with a control; and (d) determining
XX CC whether the comparison indicates a likelihood disease). (M1) is useful
XX CC for identifying modulators of the p53 pathway. A probe for HM expression
XX CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
XX CC in a patient, where the cancer has greater than 25 % expression level.
XX CC Modulators identified by (M1) are useful in a variety of diagnostic and
XX CC therapeutic applications, where disease or disorder prognosis is related
XX CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
XX CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
XX CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
XX CC the p53 function of the cell, so that the cell undergoes normal
XX CC proliferation or progression through the cell cycle. (M2) and (M3) are
XX CC also useful for treating defects in the p53 pathway such as angiogenic,
XX CC apoptotic or cell proliferation disorders. The present sequence
XX CC represents a human p53 pathway modifying protein
XX Sequence 514 AA;
XX XQ

QY 181 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSDYGFARITWKDGNLSTGLGKHGPIFALKWNKKNFILSAGVDKTTIIWDAHTGEA 300
 DB 241 ATGSDYGFARITWKDGNLSTGLGKHGPIFALKWNKKNFILSAGVDKTTIIWDAHTGEA 300
 QY 301 KQPPFHSAPALVDWQSNTPASCSTDMCIHVCKLQGDPIKTFQGHTEVNAIKWDPT 360
 DB 301 KQPPFHSAPALVDWQSNTPASCSTDMCIHVCKLQGDPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASGSDMTLKIWSMKQDNCVHDLQOHNKEIYTTIKWSPGTGNTNPNANMLASAF 420
 DB 361 GNLLASGSDMTLKIWSMKQDNCVHDLQOHNKEIYTTIKWSPGTGNTNPNANMLASAF 420
 QY 421 DSTVRLWDVRGICHTLTKHQPVPYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVRGICHTLTKHQPVPYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSY 480
 QY 481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514

RESULT 3
 ADD14051
 ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

XX 01-JAN-2004 (first entry)

DE Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;

KW protein tyrosine kinase pathway; protein tyrosine kinase; cytosolic;

KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

PN WO2003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels

CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cyrostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

Query Match 89.4%; Score 2437; DB 7; Length 577;

Best Local Similarity 86.0%; Pred. No. 4e-217;

Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLQESGFHSHTFTGKSHISQSNINGALVPPAALISIIQKGLQV 60
 DB 52 MSITDEVNFLVRYLQESGFHSHTFTGKSHISQSNINGALVPPAALISIIQKGLQV 111
 QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAA--- 117
 DB 112 EAEVISEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAAATAA 171
 QY 118 -----ASQGSAXNGENTANGAHTIANNHTDMMEDVGEIPEPNKAVVLRG 167
 DB 172 ATAATTGAGVSHQNPKNREATVNGEENRAHSV-NNHAKPMEIDGEVEIPSSKATVLRG 230
 QY 168 HSEVFIKAWNPVSDLLASGSDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKD 227
 DB 231 HSEVFIKAWNPVSDLLASGSDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKD 290
 QY 228 VTSLDWNSGTLTATGSDYGFARIWTKDGNLSTGLGKHGPIFALKWNKKNFILSAGVD 287
 DB 291 VTSLDWNTGTLTATGSDYGFARIWTKDGNLSTGLGKHGPIFALKWNKKNFILSAGVD 350
 QY 288 KTTIWDHAHTGEAKQFPFHSAPALVDWQSNTPASCSTDMCIHVCKLQGDPIKTFQ 347
 DB 351 KTTIWDHAHTGEAKQFPFHSAPALVDWQSNTPASCSTDMCIHVCKLQGDPIKTFQ 410
 QY 348 HTNEVNAIKWDPTGNLLASGSDMTLKIWSMKQDNCVHDLQOHNKEIYTTIKWSPGTG 407
 DB 411 HTNEVNAIKWDPTGNLLASGSDMTLKIWSMKQDNCVHDLQOHNKEIYTTIKWSPGTG 470
 QY 408 NPNANMLASAFDSTVRLWDVRGICHTLTKHQPVPYVAFSPDGRYLAGSFDKCVH 467
 DB 471 NPNANMLASAFDSTVRLWDVRGICHTLTKHQPVPYVAFSPDGRYLAGSFDKCVH 530
 QY 468 IWNTOGALVHSYRGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
 DB 531 IWNTOGALVHSYRGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 577

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX ABP41760;

XX 22-AUG-2002 (first entry)

DE Human ovarian antigen HNOKM38, SEQ ID NO:2892.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 OS Homo sapiens.
 XX
 XX WO200200677-A1.
 PN
 XX
 XX 03-JAN-2002.
 PD
 XX
 XX 07-JUN-2001; 2001WO-US018569.
 PF
 XX
 XX 07-JUN-2000; 2000US-0209467P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Birse CE, Rosen CA;
 PI
 XX
 XX WPI, 2002-147878/19.
 DR
 XX
 XX N-PSDB; ABQ54837.

Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 cancer), immune disorders, cardiovascular disorders and neurological
 diseases.

Claim 11; SEQ ID NO 2892; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-
 ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 encompasses polypeptides 90% identical and polynucleotides 95% identical
 to the sequences of the invention. The invention additionally relates to
 recombinant vectors and host cells comprising human ovarian antigen
 polynucleotides, antibodies against human ovarian antigens, and the use
 of ovarian antigen polynucleotides and polypeptides in diagnosing,
 treating, prognosing or preventing various ovary and/or breast-related
 disorders. Such conditions include ovarian cancer and breast cancer, and
 metastatic tumours of ovarian or breast origin, reproductive system
 disorders (e.g., infertility, disorders of pregnancy, anovulation,
 polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 vaginitis), immune disorders (e.g., congenital and acquired
 immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 blood-related disorders (e.g., anaemia), cardiovascular disorders,
 respiratory disorders, neurological disorders, gastrointestinal disorders
 and urinary system disorders. Ovarian antigen polypeptides and
 polynucleotides may also be used in screening for compounds which
 modulate ovarian antigen expression or activity. The polynucleotides may
 further be used for gene therapy, chromosome mapping, in the
 identification of individuals and in forensic analysis, and the
 polypeptides may be used as food additives or to prepare antibodies
 useful in disease diagnosis, drug targeting and phenotyping. The present
 sequence represents a human ovarian antigen of the invention. Note: The
 sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences

Sequence 542 AA;

Query Match 85.8%; Score 2340; DB 5; Length 542;
 Best Local Similarity 85.4%; Fred. No. 3.8e-208;
 Matches 434; Conservative 32; Mismatches 28; Indels 14; Gaps 2;
 20 GFSHAFTFGKSHISQSNINGALVPPAALISIIKQLQYVEAEVSEINEDGTLDGRPIE 79
 36 GFSHAFTFXLESHISQSNINGTLVPPAALISIIKQLQYVEAEISINEDGTVDGRPIE 95

QY 80 SLSLIDAVMPDVVQTRQQAAYRDKLAQQQAAAAA-----ASQCGSAXN 126
 DB 96 SLSLIDAVMPDVVQTRQQAAYRDKLAQQQAAAAA-----AAATAATTTTTSAGVSHQNSKN 155
 QY 127 GENTANGENGAAHTTANNHTDMVGDGVEIPPNKAVLVLRGHSESEVFICANNPYSDLLAS 186
 DB 156 REATVNGEENRAHSV-NNHAKPMEIDGVEIIPSSKATVLRGHSESEVFICANNPYSDLLAS 214
 QY 187 GSGDSTARLWNLSENSTGSLVLRHCIRREGQDVPNSKDVTSIDMNSSEGLLATGSYD 246
 DB 215 GSGDSTARLWNLSENSTGSLVLRHCIRREGQDVPNSKDVTSIDMNSSEGLLATGSYD 274
 QY 247 GFARIWTKDGNLASTLGQHKGPFFALKKNKKNFILSAGVDKTTIINDAHTGEAKQPPFF 306
 DB 275 GFARIWTKDGNLASTLGQHKGPFFALKKNKKNFILSAGVDKTTIINDAHTGEAKQPPFF 334
 QY 307 HSAPALDWDQSNNTFASGSTDWCIVHCKIGDRPIKTFQGHNEVNNAIKWDPCTGNLLAS 366
 DB 335 HSAPALDWDQSNNTFASGSTDWCIVHCKIGDRPIKTFQGHNEVNNAIKWDPCTGNLLAS 394
 QY 367 CSDDMTLKIWSMKQDNCVHDLLQHNKEIYTIKWSPTGPTNNPNANMLASGSDSTVRL 426
 DB 395 CSDDMTLKIWSMKQDNCVHDLLQHNKEIYTIKWSPTGPTNNPNANMLASGSDSTVRL 454
 QY 427 WVDVDRGICHTLTIKHQPYPYVAFSPDGRYLASGSDFKCVHNTQTGALVHSYRGTTGI 486
 DB 455 WVDVDRGICHTLTIKHQPYPYVAFSPDGRYLASGSDFKCVHNTQTGALVHSYRGTTGI 514
 QY 487 FEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 515 FEVCNNAAGDKVGASASDGSVCVLDLRK 542
 RESULT 5
 ABP51424
 ID ABP51424 standard; protein; 395 AA.
 XX
 AC ABP51424;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 446.
 XX
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiaslathmatic; antiatherosclerotic; antigout;
 KW neuroprotective; antirheumatic; antiarthritic.
 XX
 OS Homo sapiens.
 XX
 XX WO200240715-A2.
 PN
 XX
 PD 23-MAY-2002.
 XX
 XX 06-SEP-2001; 2001WO-US027628.
 PF
 XX
 XX 05-SEP-2000; 2000US-0229747P.
 PR
 XX 05-SEP-2000; 2000US-0229748P.
 PR
 XX 05-SEP-2000; 2000US-0229749P.
 PR
 XX 05-SEP-2000; 2000US-0229750P.
 PR
 XX 05-SEP-2000; 2000US-0229751P.
 PR
 XX 05-SEP-2000; 2000US-0230583P.
 PR
 XX 06-SEP-2000; 2000US-0230505P.
 PR
 XX 06-SEP-2000; 2000US-0230514P.
 PR
 XX 06-SEP-2000; 2000US-0230515P.
 PR
 XX 06-SEP-2000; 2000US-0230517P.
 PR
 XX 06-SEP-2000; 2000US-0230518P.
 PR
 XX 06-SEP-2000; 2000US-0230519P.
 PR
 XX 06-SEP-2000; 2000US-0230595P.
 PR
 XX 06-SEP-2000; 2000US-0230597P.


```

PR 06-SEP-2000; 2000US-0230598P.
PR 06-SEP-2000; 2000US-0230599P.
PR 06-SEP-2000; 2000US-0230610P.
PR 06-SEP-2000; 2000US-0230612P.
PR 06-SEP-2000; 2000US-0230663P.
PR 06-SEP-2000; 2000US-0230988P.
PR 06-SEP-2000; 2000US-0230989P.
PR 07-SEP-2000; 2000US-0230951P.
PR 07-SEP-2000; 2000US-0231163P.
PR 07-SEP-2000; 2000US-0231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
XX Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
XX Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
XX Roseberry AM, Garstin EH, Peralta CH, David MH, Panzer SR, Flores V;
XX Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;
XX WPI; 2002-527544/56.
XX N-PSDB; ABQ72641.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
XX AIDS.
XX
XX Claim 14; Page 578; 618pp; English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDDT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDDT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDDT. (I)
XX or (II) are useful for diagnosing, treating or preventing disorders
XX associated with aberrant expression of MDDT, where the disorders are
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germline gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences.
XX
XX Sequence 395 AA;
XX
XX Query Match
XX Best Local Similarity 75.1%; Score 2047; DB 5; Length 395;
XX Matches 389; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MSISDEVNFLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQGLQYV 60
XX
XX 5 MSISDEVNFLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQGLQYV 64
XX
XX 61 EAEVISEDGTLFDRPTESLSLDVMPDVVQTEQQAAYRDKLACQAAAAAAAASQ 120
XX
XX 65 EAEVISEDGTLFDRPTESLSLDVMPDVVQTEQQAAYRDKLACQAAAAAAAASQ 124

```

```

QY 121 QGSAKNGENTANGENGCAHTIANNHTDMMVDGVEIPPNKAVVLRGHESEVFI CAWN PV 180
DB 125 QGSAKNGENTANGENGCAHTIANNHTDMMVDGVEIPPNKAVVLRGHESEVFI CAWN PV 184
QY 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB 185 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
QY 241 ATGSYDGFARINTKDGNLASTLQGHKGPIPALKNKKNKGNFILSAGVDKTTIIIDAHNTGEA 300
DB 245 ATGSYDGFARINTKDGNLASTLQGHKGPIPALKNKKNKGNFILSAGVDKTTIIIDAHNTGEA 304
QY 301 KQFPFPHSAFALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNTNEVNAIKWDPT 360
DB 305 KQFPFPHSAFALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNTNEVNAIKWDPT 364
QY 361 GNLLASCSDDMTLKIWSMKQDNVCVHDLQAHN 391
DB 365 GNLLASCSDDMTLKIWSMKQDNVCVHDLQAHN 395

```

RESULT 6

ABG21351
ID ABG21351 standard; protein; 584 AA.

XX AC ABG21351;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #21342.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX XX W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS85538.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX PS Claim 20; SEQ ID NO 51710; 103pp; English.

XX SS The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in

QY 353 NAIKWDPDPTNLLASCSDDMTLKIMSKQDNCVHDLQHNKEIYTIKWSPTGPGTNNPNAN 412
 DB 538 NAIKWCPQOQLASCSDDMTLKIMSMNRDRCHDLQAHSEIYTIKWSPTGPGTNNPNTN 597
 QY 413 LMLASASDSTVRLWDVDRGICIIHTLTKHQBPVYVSAFSPDGRVLAGSPKCVHIWNTQ 472
 DB 598 LILASASDSTVRLWDVDRGSCIIHTLTKHTEFPVYVSAFSPDGKHLASGSPKCVHIWSTQ 657
 QY 473 TQALVHSYRGTGGIFEVWCNAGDKVGSASDGSVCVLDLRK 514
 DB 658 TQGLVHSYKGTGGIFEVWCNSKGTKVGSASDGSVFLDLRK 699

RESULT 8
 AAO04385
 ID AAO04385 standard; protein; 208 AA.
 XX AC AAO04385;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 18277.
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX OS Homo sapiens.
 XX FN WO200164835-A2.
 XX PD 07-SEP-2001.
 XX PF 26-FEB-2001; 2001WO-US004927.
 XX PR 28-FEB-2000; 2000US-0051526.
 XX PR 18-MAY-2000; 2000US-00577409.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX DR WPI; 2001-514838/56.
 XX DR N-PSDB; AAI84316.
 XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 XX and treating e.g. leukemia, inflammation and immune disorders.
 XX PS Claim 20; SEQ ID NO 18277; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 208 AA;
 Query Match 41.7%; Score 1137; DB 4; Length 208;
 Best Local Similarity 99.5%; Pred. No. 5.5e-97;
 Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 213 HCIREGGQDVPSKDVTSLDWNSRGTLATGSDGFARITWKDGNLASTLGQHKGPITAL 272
 |||||

DB 1 HCIREGGQDVPSKDVTSLDWNSRGTLATGSDGFARITWKDGNLASTLGQHKGPITAL 60
 QY 273 KNNKGNFILSAGVDKTTIIWDAHTGEAKQPPFHSAPALDWDQSNNTFASGSTDGCIH 332
 DB 61 KNNKGNFILSAGVDKTTIIWDAHTGEAKQPPFHSAPALDWDQSNNTFASGSTDGCIH 120
 QY 333 VCKLGQDRPIKTFQGHTEVNAIKWPTGNTLILASCSDDMTLKIMSKQDNCVHDLQHNK 392
 DB 121 VCKLGQDRPIKTFQGHTEVNAIKWPTGNTLILASCSDDMTLKIMSKQDNCVHDLQHNK 180
 QY 393 EIYTIKWSPTGPGTNNPNANLMLASAS 419
 DB 181 EIYTIKWSPTGPGTNNPNANLMLASAS 207

RESULT 9
 ABG21350
 ID ABG21350 standard; protein; 167 AA.
 XX AC ABG21350;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #21341.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX FN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS85537.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.

XX Claim 20; SEQ ID NO 51709; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at

CC	ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 167 AA;	
SQ		
Query Match	30.0%; Score 819; DB 4; Length 167;	
Best Local Similarity	86.2%; Pred. No. 1.4e-67;	
Matches 144; Conservative	13; Mismatches 10; Indels 0; Gaps 0;	
QY	250 RIWTDGNLSTGLGKGPFIKALWKKGNFILSAGVDKTTIWDATGEAKQFPFSA 309	
DB	1 RIWTDGNLSTGLGKGPFIKALWKKGNFILSAGVDKTTIWDATGEAKQFPFSA 60	
QY	310 PALVDWQNSNTFASCSTDMCIHVCKLQDRIPIKTFQHTNEVNAIKWDPTGNLLASCSD 369	
DB	61 PALVDWQNSNTFASCSTDMCIHVCKLQDRIPIKTFQHTNEVNAIKWDPTGNLLASCSD 120	
QY	370 DMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLA 416	
DB	121 DMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPTGNPNANMLA 167	
RESULT 10		
ID	ABR52980 standard; protein; 535 AA.	
XX	ABR52980;	
AC	ABR52980;	
DT	20-JUN-2003 (first entry)	
XX		
DE	Protein sequence #SEQ ID 825.	
XX		
KW	Multiprotein complex; eukaryote; drug target; diagnosis.	
XX		
OS	Saccharomyces cerevisiae.	
XX		
PN	EP1258494-A1.	
XX		
PD	20-NOV-2002.	
XX		
PF	20-DEC-2001; 2001EP-00130253.	
XX		
PR	15-MAY-2001; 2001EP-00111774.	
XX		
PA	(CELL-) CELLZONE AG.	
XX		
PI	Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;	
PI	Marzioch M, Schultz JD, Superti-Furga GD;	
XX		
DR	WPI; 2003-250078/25.	
DR	N-PSDB; ACC61022.	
XX		
PT	New isolated protein complexes useful for diagnosing a disease or	
PT	disorder, or as a target for an active agent of a pharmaceutical,	
PT	preferably a drug target in the treatment or prevention of disease or	
PT	disorder.	
PS	Disclosure; SEQ ID NO 825; 17pp + Sequence Listing; English.	
XX		
CC	The invention relates to multiprotein complexes from eukaryotes. Proteins	
CC	of the invention and DNA sequences encoding them are given in records	
CC	of the invention and ACC60610-ACC61944 respectively. The complexes are	
CC	obtainable by using a protein as a bait and isolating the set of proteins	
CC	which is attached thereto from cells. Such protein complexes may comprise	
CC	up to 30 distinct proteins. Protein complexes of the invention are useful	
CC	for diagnosing a disease or disorder, or as a target for an active agent	
CC	of a pharmaceutical, preferably a drug target in the treatment or	
CC	prevention of a disease or disorder. Note: The sequence data for this	
CC	patent is not represented in the printed specification, but is based on	
CC	sequence information supplied by the European Patent Office. The complete	
CC	document is available on CD-ROM	
XX		
SQ	Sequence 535 AA;	

Query Match	18.3%; Score 499.5; DB 6; Length 535;	
Best Local Similarity	24.9%; Pred. No. 4.3e-37;	
Matches 142; Conservative	100; Mismatches 232; Indels 97; Gaps 15;	
QY	1 MSISSDEVNFLVRYLQSGFSGHSA-----FTFGIKSHISQSNINGALVPPAAL 49	
DB	1 MSITSEELNYLWRYCOEMGHEVSALALQDETRVLEFDEKYEKHI-----PLGTL 50	
QY	50 ISIIQGLQYVAEVSINEDG---TLFQGRPIESLSLDAMVPDVVQTRQAYRKLAAQ 106	
DB	51 VNLVQRGILYTESLMDKSGDISALNEHLSDEFNLVQALQID-----KEKFEI 101	
QY	107 QAAAAAASQASAKNGENTANGENGAAHTIANHNTDMVEYDGVVEIPPNKAVLVR 166	
DB	102 SEBGRFTLETNESKAGDGASTVERTOEDDTNSIDSSD---DLDDGFVKI---LKEIV-- 155	
QY	167 GHESEVFIICAMPVSD--LLASGSGDSTARIMWLSNENSTSGSTQ-----LVLRHCIREGG 219	
DB	156 -KLDNIVSSTWNPLOESILAYGEKNSVARLARIVETDQEGKYYKLTITIAELRHPFALSA 214	
QY	220 QDVPSNKDVTSLDNNSEGLTLLATGSDGFARITWTKDGNLSTGLGKGPFIKALWKKGN 279	
DB	215 SSGKTTNQVTCIAWSHDGNSIVTGVENGELRWKNTGALLNVLNFRAPIVSVKANKDGT 274	
QY	280 FILSAGVDKTTIWDATGEAKQFPF-----HSAP---ALDVMQSNNTFAS 324	
DB	275 HIISMDVENVTILWNVISGVVQHPKELKGTGSSNAENHSGDGLGVDEWVDDDKFYI 334	
QY	325 CSTDMCIHVCKLQDRIPIKTFQHTNEVNAIKWDPTGNLLASCSDMTLKWSMKQDNCV 384	
DB	335 PGPKGAIFYQITKTEPTCKLIGHHGPISVLFBNFTNKLKLLSASDDGTURIWHGNGNSQ 394	
QY	385 HDLQHNKEIYTIKWSPTGPTGNPNANMLASAFSDSTVRLWDVDRGICHTLTKHOEP 444	
DB	395 NCFYHGSQISVSNWV-----GDDKVISCSWDGSGVRLWSLKNQTLALSISIVDGV 444	
QY	445 VYSVAFSPDGRYIASGSPKCVHIW-----NTQTGAL-----VHSYR 481	
DB	445 IFAGRISQDQKYAVAFMDGVQVNVYDLKLNKSKRSLYGNRDLNPLPIPIYASYQSSQ 504	
QY	482 GTGGIEVCWNAAGDKVGA--SASDGSVCVL 510	
DB	505 DNDYIFDLSWNCAGNKISVAYSLOEGSVVAI 535	
RESULT 11		
ID	ADA13321 standard; protein; 414 AA.	
XX	ADA13321;	
AC	ADA13321;	
XX		
DT	06-NOV-2003 (first entry)	
XX		
DE	Human intracellular signalling molecule INTS1G-3, SEQ ID NO:3.	
XX		
KW	Human; intracellular signalling molecule; INTS1G;	
KW	cell proliferative disorder; cancer; atherosclerosis;	
KW	autoimmune disorder; inflammatory disorder; infection;	
KW	neurological disorder; developmental disorder; endocrine disorder;	
KW	cytostatic; antiarteriosclerotic; nootropic; neuroprotective;	
KW	cerebroprotective; immunosuppressive; antiinflammatory; gene therapy.	
OS	Homo sapiens.	
XX		
FN	W02003031568-A2.	
XX		
PD	17-APR-2003.	
XX		
PF	16-AUG-2002; 2002WO-US026322.	
XX		
PR	17-AUG-2001; 2001US-0313245P.	
PR	24-AUG-2001; 2001US-0314751P.	
PR	31-AUG-2001; 2001US-0316752P.	

Db 168 RLASCKAGSIIMPETGQKGRPLSGHKHINCLAWEPVHRDPECKKLASAGDGD 227
 QY 251 IW-TKDGNLASTLGHKGPFIKLNKKNKNGILSAGVDKTIWIIDAHTGEAKQFPFSA 309
 Db 228 IWDVGLGQCLMNIAGHTNAVTAVRWGGAG-LIYTSSKDRIVKMWRAADGILCRITFSGHA- 285
 QY 310 PALDVEDWQNNTFASCSDTMCIHVCKLGDRIK-TFOGH-----TNEV-----NAIK 356
 Db 286 -----HWNN-----IALSTD-----YVLTGTFPHKDRSKSHLSLSTEELQESALKRYQAVC 334
 QY 357 WDPTGNLLASDDMTLKIWSKQNCVHDLQHNKEIYTIKWSPGTGPTNNPNNMLLA 416
 Db 335 PDEVESLV-SCSDNTLYLRNNQNKVCVERMTGHQNVVNDVKYSPDVK-----LIA 384
 QY 417 SASFDSTVRLNDVDRGICHTLTKEQEPVYSVAFSPDGRYLASGDFDKCVHIWNTQTGAL 476
 Db 385 SASFDKSVRLWRASDQYMATFRGHVQAVYIVAMSADSLIVSGKSTLKWVSQTKL 444
 QY 477 VHSYRG-TGGIFECVNAAGDKVGASD 504
 Db 445 AQELPGHAEVFGVDWAPDGSRVASGGKD 473

RESULT 13

AA79678
 ID AA79678 standard; protein; 479 AA.

XX AC AA79678;
 XX DT 29-AUG-2000 (first entry)
 XX DE Drosophila Notchless protein.
 XX KW Notchless; Nle Gene; Notch; signalling; neurodegenerative disease;
 KW cancer; diagnosis; cytostatic; neuroprotective; therapy.
 XX OS Drosophila melanogaster.
 XX FH Key Location/Qualifiers
 FT Domain 27..105
 FT /note= "Nle domain"
 FT Misc-difference 67
 FT /note= "sequence deduced from nucleotide sequence has an
 FT additional lys residue between Lys-67 and Ser-68"
 FT Misc-difference 282
 FT /note= "encoded by CAA"

WO200026364-A1.

PD 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB001891.

XX 03-NOV-1998; 98GB-00024045.

XX (EUWO-) EURO MOLECULAR BIOLOGY LAB.

XX Cohen S, Boumeester A, Royet J;

XX WPI; 2000-365613/31.

XX N-PSDB; AAA27739.

PT Novel notchless protein and nucleic acids encoding them useful for
 PT treating and preventing cancer and neurodegenerative diseases.

XX Claim 1; Page 44; 52pp; English.

XX The present sequence is that of Notchless, a novel protein of Drosophila.
 CC Notchless was identified in a screen for dominant modifiers of a Notch
 CC mutant phenotype in the Drosophila wing. The mutant dominantly suppressed
 CC the wing notching phenotype of notchoid mutations, and the Notchless
 CC protein was shown to bind to the cytoplasmic domain of Notch. Notchless

CC modified Notch signalling activity in a variety of Notch-dependent
 CC signalling process in both Drosophila and Xenopus embryos. The Notchless
 CC protein has a novel highly conserved N-terminal domain followed by 9 WD40
 CC repeats. Notchless, and nucleic acids encoding it, can be used in methods
 CC for the diagnosis and therapy of certain diseases, particularly cancer
 CC and neurodegenerative diseases (claimed). A Notchless mutant in a
 CC sensitised Notch genetic background is used in a claimed method for
 CC identifying compounds capable of modifying the levels of expression or
 CC activity of a Notch protein

XX Sequence 479 AA;

Query Match 13.6%; Score 370; DB 3; Length 479;

Best Local Similarity 28.4%; Pred. No. 4e-25;

Matches 120; Conservative 55; Mismatches 158; Indels 90; Gaps 18;

QY 147 DMMEVDG---DVEIIPNKAVVLR-----GHESEVFICAWNPFVSDLLASGSGDST 192
 Db 74 DLASVDTENVIDIVYQPAVFRVTRCTSSMPGHAEAVSLNFSPOGAHLASGSGDTT 133
 QY 193 ARIWNLSEN-----STSGSTQLVLRHCIR-----EGQD----- 221
 Db 134 VRLWDLNTETPHCTCTGHKQWVL--CVSWAPDGKRLASGCKAGSIIIWDPETGQCKGRPL 191
 QY 222 VPSNKDVTSLDW-----NSEGTLATGSGYDGFARIW-TKDGNLASTLGHKGPFIKLNKWN 275
 Db 192 SGHKHINCLAWEPVHRDPECKKLASAGDGDRIWVGLGQCLMNIAGHTNAVTAVRWG 251
 QY 276 KKGNFILSAGVDKTIWIIDAHTGEAKQFPFHSAPALDVEDWQNNTFASCSDTMCIHVCK 335
 Db 252 GAG-LIYTSSKDRIVKMWRAADGILCRITFSGHA-----HWNN-----IALSTD-----YVLR 298
 QY 336 LGQDRPIK-TFOGH-----TNEV-----NAIKWPTGNLLASGSDMTLKIWSKQDN 382
 Db 299 TGFPHKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLV-SCSDNTLYLRNNQNK 357
 QY 383 CVHDLQHNKEIYTIKWSPGTGPTNNPNNMLLASASFDSTVRLNDVDRGICHTLTKHQ 442
 Db 358 CVERMTGHQNVVNDVKYSPDVK-----LIAASFDKSVRLWRASDQYMATFRGHV 408
 QY 443 EPVYSVAFSPDGRYLASGDFDKCVHIWNTQTGALVHSYRG-TGGIFECVNAAGDKVGAS 501
 Db 409 QAVYIVAMSADSLIVSGKSTLKWVSQTKLAQELPGHAEVFGVDWAPDGSRVASG 468
 QY 502 ASD 504
 Db 469 GKD 471

RESULT 14

ABB62260
 ID ABB62260 standard; protein; 411 AA.

XX AC ABB62260;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 13572.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical.

XX OS Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 2726

Sequence: 1 MSISDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2716	99.6	514	1	TBLR_HUMAN
2	2437	89.4	526	1	Q9BZK7: O9H211; Q9H9A1;
3	2367	86.8	522	1	TELX_HUMAN
4	1304	47.8	313	1	TELX_MOUSE
5	502.5	18.4	535	1	SIF2 YEAST
6	479	17.6	1526	1	Q8YR11 anabaena sp
7	463.5	17.0	1258	1	Y500 ANASP
8	460	16.9	1356	1	Q00808 podotheca a
9	445.5	16.3	1683	1	Q8YV57 anabaena sp
10	388.5	14.3	1693	1	Y163 SYN3
11	366.5	13.4	485	1	HUS7_HUMAN
12	365.5	13.4	485	1	YCW2 YEAST
13	365.5	13.4	515	1	AAC3_DICDI
14	348.5	12.8	437	1	CIAL_HUMAN
15	348	12.8	339	1	L181_HUMAN
16	346.5	12.7	409	1	L181_BOVIN
17	345.5	12.6	409	1	L181_MOUSE
18	344.5	12.6	409	1	WDS DROME
19	340.5	12.5	361	1	WDR5_HUMAN
20	339.5	12.5	334	1	T2D4_HUMAN
21	339	12.4	800	1	PKWA_THECU
22	334	12.3	742	1	T2D4_YEAST
23	329	12.1	798	1	Y143 SYN3
24	321	11.8	376	1	Y143 SYN3
25	319.5	11.7	606	1	PF20 CHLRE
26	317.5	11.6	579	1	SE10_CABEL
27	310.5	11.4	704	1	T2D4_DROME
28	309	11.3	1191	1	Y143 SYN3
29	308	11.3	640	1	MT30_YEAST
30	307	11.3	678	1	SCOB_TMENTI
31	304	11.2	605	1	POF1_SCHPO
32	300	11.0	665	1	L123_CABEL
33	294.5	10.8	501	1	YH92_CABEL

ALIGNMENTS

RESULT 1

```

TBLR_HUMAN
ID TBLR_HUMAN STANDARD; PRT; 514 AA.
AC Q9BZK7: O9H211; Q9H9A1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBL1-
DE related protein 1).
GN TBLR1 OR IRA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.
RX MEDLINE=21930350; PubMed=11931768;
RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;
RT "The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK
RT pathway through the integral subunit GPS2.";
RL Mol. Cell 9:611-623(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang X., Dormady S., Basch R.;
RT "Identification of four human cDNAs that are differentially expressed
RT by early hematopoietic progenitors.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: Interacts with HDAC3.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF314544; AAK03001.1; -
CC EMBL; AF268193; AAG44736.1; -
CC EMBL; AK022956; BAB14331.1; -
CC InterPro; IPR006594; Lish.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 8.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD0000018; WD40; 3.
CC SMART; SM00667; Lish; 1.

```

```

34 294.5 10.8 643 1 T2D4_SCHPO O13282 schizosacch
35 292.5 10.7 614 1 TUL1_SCHPO Q09715 schizosacch
36 291 10.7 395 1 YZLL_CAEEL Q93847 caenorhabdi
37 289.5 10.6 395 1 GBB5_HUMAN O14775 homo sapien
38 287.5 10.5 395 1 GBB5_MOUSE P54314 mus musculu
39 285 10.5 586 1 TUL2_SCHPO Q9uug8 schizosacch
40 283 10.4 605 1 FW1A_HUMAN O9v297 homo sapien
41 283 10.4 931 1 LUG_ZEATH O9fuy2 arabidopsi
42 281.5 10.3 554 1 PRP4_ARATH O22212 arabidopsi
43 280.5 10.3 380 1 GBB_ZAIZE P49178 zea mays (m
44 280.5 10.3 713 1 TUP1_YEAST P16649 saccharomyc
45 279 10.2 522 1 PRP4_HUMAN O43172 homo sapien

```

DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 167 206 WD 1.
 FT REPEAT 223 262 WD 2.
 FT REPEAT 264 303 WD 3.
 FT REPEAT 306 344 WD 4.
 FT REPEAT 347 386 WD 5.
 FT REPEAT 389 437 WD 6.
 FT REPEAT 440 479 WD 7.
 FT REPEAT 481 513 WD 8.
 FT DOMAIN 108 118 POLY-ALA.
 FT CONFLICT 31 31 E -> K (IN REF. 2).
 FT CONFLICT 59 59 Y -> H (IN REF. 3).
 FT CONFLICT 389 389 A -> Q (IN REF. 2).
 SQ SEQUENCE 514 AA; 55595 MW; 08556D2E4BA796D CRC64;
 Query Match 99.6%; Score 2716; DB 1; Length 514;
 Best Local Similarity 99.6%; Pred. No. 1.5e-188;
 Matches 512; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSISDEVNFLVRYLQBSGFSHAFTGKISHISQSNINGALVPPAALISITQKGLQYV 60
 DB 1 MSISDEVNFLVRYLQBSGFSHAFTGIESHISQSNINGALVPPAALISITQKGLQYV 60
 QY 61 EAEVSNEDGTLFQGRPIESLSLDVMPDVQVQYQAYRDKLAQQAQAAAAAASQ 120
 DB 61 EAEVSNEDGTLFQGRPIESLSLDVMPDVQVQYQAYRDKLAQQAQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHDMVEVDGVEIPPKAVVLRGHSEVFICANVPV 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHDMVEVDGVEIPPKAVVLRGHSEVFICANVPV 180
 QY 181 SDLLASGSGDSTARVNLSENSTSGTQVLVLRHCIRGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SDLLASGSGDSTARVNLSENSTSGTQVLVLRHCIRGGQDVPSNKDVTSLDWNSEGTLL 240
 QY 241 ATGSYDGFARITWQGNLSTLQGHKGFIFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSYDGFARITWQGNLSTLQGHKGFIFALKWKKGNFILSAGVDKTTIINDAHTGEA 300
 QY 301 KOQPFHSAPALVDVQSNNTFASCTDMCHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 DB 301 KOQPFHSAPALVDVQSNNTFASCTDMCHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQNCVVDHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQNCVVDHDLQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLMDVDRGICHTLTKQBPVYVAFSPDGRYLASGDFKCVHINTQTALVHSY 480
 DB 421 DSTVRLMDVDRGICHTLTKQBPVYVAFSPDGRYLASGDFKCVHINTQTALVHSY 480
 QY 481 RGTGGIFEVCVNNAAGDKVGASDGSCVCLDLRK 514
 DB 481 RGTGGIFEVCVNNAAGDKVGASDGSCVCLDLRK 514
 RESULT 2
 ID TLBX HUMAN
 AC 060907;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Transducin beta-like 1X protein (Transducin-beta-like 1, X-linked).
 GN TLBX OR TBL1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 [1] NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99264241; PubMed=10330347;
 RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
 Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;
 RT "X-linked late-onset sensorineural deafness caused by a deletion
 involving OAI and a novel gene containing WD-40 repeats.",
 RL Am. J. Hum. Genet. 64:1604-1616(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Lymph;
 MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smal M.A.,
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Contains 1 Lish domain.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 EMBL; Y12781; CAA73319.1; ALT_INIT.
 EMBL; BC032708; AAH32708.1; -.
 Genew; HGNC:11585; TBL1X.
 MIM; 300196; -.
 GO; GO:0007605; P:hearing; TAS.
 GO; GO:0007165; P:signal transduction; TAS.
 GO; GO:0007601; P:vision; TAS.
 GO; GO:0007601; P:vision; TAS.
 InterPro; IPR006594; Lish.
 InterPro; IPR001860; WD40.
 Pfam; PF04000; WD40; 8.
 PRINTS; PR00320; GPROTEINRPT.
 ProDom; PD000018; WD40; 3.
 SMART; SM00667; Lish; 1.
 SMART; SM00320; WD40; 7.
 PROSITE; PS00896; LISH; 1.
 PROSITE; PS00678; WD_REPEATS_1; 4.
 PROSITE; PS00082; WD_REPEATS_2; 6.
 PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT DOMAIN 4 36 LISH.
 FT REPEAT 167 206 WD 1.
 FT REPEAT 223 262 WD 2.
 FT REPEAT 264 303 WD 3.
 FT REPEAT 306 344 WD 4.
 FT REPEAT 347 386 WD 5.
 FT REPEAT 389 437 WD 6.
 FT REPEAT 440 479 WD 7.
 FT REPEAT 481 513 WD 7.

```
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 POLY-ALA.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match
Best Local Similarity 89.4%; Score 2437; DB 1; Length 526;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNLFVRYLQSGSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSITSDVNLFVRYLQSGSHSAFTFGIESHISQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVINEDEGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 117
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 EAEISINEDEGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 120
QY 118 -----ASQGSAXNGENTANGEAHTIANNHTDMVEVDGVEIIPNKAVLVRG 167
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ATAATTAGVSHQNPKNREATVNGEENRAHSV-NNHAKPMEIDGVEIIPSKATVLRG 179
QY 168 HSEVFIICAWNPVSDLLASGSGDSTARIWNLSNSTGSTQVLVRHCIREGGQVPSNKG 227
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 HSEVFIICAWNPVSDLLASGSGDSTARIWNLSNENSGSTQVLVRHCIREGGHVDVPSNKG 239
QY 228 VTSLDWNSGTLATGSDGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVD 287
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 VTSLDWNTGTLATGSDGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVD 299
QY 288 KTTIWDHTGEAKQFPFHSAAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQG 347
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 KTTIWDHTGEAKQFPFHSAAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQG 359
QY 348 HTNEVNAIKWPTGNLLASCDDMTLKIWSKQDNCVHDLOHKNKEIYTIKWSPTGPGTN 407
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 HTNEVNAIKWPSGMLLASCDDMTLKIWSKQEVCHIDLQAHKEIYTIKWSPTGATS 419
QY 408 NPANMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 467
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 NPNSIMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 468 IWTQTGALVHSYRGTTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 IWTQSGNLVHSYRGTTGGIFEVCAAGDKVGASASDGSVCVLDLRK 526

RESULT 3
ID TBLY_HUMAN
AC Q9BQ87.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBLY OR TBLY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```

or send an email to license@isb-sib.ch).

```
CC EMBL; AF332220; AAK13472.1; -
CC EMBL; AF332221; AAK13473.1; -
DR EMBL; AF332222; AAK13474.1; -
DR Genew; HGNC:18502; TBLY.
DR InterPro; IPR006594; LISH.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match
Best Local Similarity 86.8%; Score 2367; DB 1; Length 522;
Matches 442; Conservative 35; Mismatches 34; Indels 12; Gaps 2;

QY 1 MSISDEVNLFVRYLQSGSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSITSDVNLFVRYLQSGSHSAFTFGIESHISQSNINGTLVPPAALISIIQKGLQYV 60
QY 61 EAEVINEDEGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 115
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 EAEISINEDEGTLFDGRPIESLSIIDAVMPDVVQTRQAVRDKLAQOQAAAAA--- 120
QY 116 -----AASQGSAXNGENTANGEAHTIANNHTDMVEVDGVEIIPNKAVLVRGHE 169
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 AATMTAAISQNPKNREATVNGEENGAHEI-NNHAKPMEIDGVEIIPNKATVLRGHE 179
QY 170 SEVFIICAWNPVSDLLASGSGDSTARIWNLSNSTGSTQVLVRHCIREGGQVPSNKGDT 229
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 SEVFIICAWNPVSDLLASGSGDSTARIWNLSNENSGSTQVLVRHCIREGGHVDVPSNKGDT 239
QY 230 SLDWNSGTLATGSDGFARITWDGNLSTLQGHKGFIFALKWNNKGNFILLSAGVDKT 289
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 SLDWNSDGTLLAMGSDGFARITWENGLASTLQGHKGFIFALKWNNKGNVILLSAGVDKT 299
QY 290 TTIWDHTGEAKQFPFHSAAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQHT 349
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 TTIWDHTGEAKQFPFHSAAPALDVQSNNTFASCSTDMCHIVCKLGDRIKTFQHT 359
QY 350 NEVNAIKWPTGNLLASCDDMTLKIWSKQDNCVHDLOHKNKEIYTIKWSPTGPGTN 409
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 NEVNAIKWPSGMLLASCDDMTLKIWSKQDNCVHDLOHKNKEIYTIKWSPTGATS 419
QY 410 NANMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 469
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 NSSIMLASAFSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 470 NTQTGALVHSYRGTTGGIFEVCAAGDKVGASASDGSVCVLDL 512
Dd |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 NTQSGSLVHSYRGTTGGIFEVCAAGDKVGASASDGSVCVLDL 522

RESULT 4
ID TBLY_MOUSE
PRT; 313 AA.
STANDARD;
```



```
DR PROSITE; PS50082; WD REPEATS 2; 4.
KW Repeat; WD repeat; Chromatin_regulator; Nuclear protein.
FT DOMAIN 4 36
FT REPEAT 163 208
FT REPEAT 218 257
FT REPEAT 259 298
FT REPEAT 315 354
FT REPEAT 396 437
FT REPEAT 440 479
FT REPEAT 496 535
FT CONFLICT 396 437
SQ SEQUENCE 535 AA; 59145 MW; 25DD19AB2FB4B07 CRC64;

Query Match 18.4%; Score 502.5; DB 1; Length 535;
Best Local Similarity 24.9%; Pred. No. 8.2e-29;
Matches 142; Conservative 100; Mismatches 232; Indels 97; Gaps 15;

Qy 1 MSISDEVNFLVYRLQESGFSEA-----TFGIKSHISQSNINGALVPPAAL 49
Db 1 MSITSEELNYLIWRYCOEMGHEVSALALQDTRVLEFDEKYKEHI-----PLGTL 50

50 ISITQKGLQYVEAEVSNEDG---TLFGRPIESLSLIDAVPDVWQTRQAVKDKLAQQ 106
51 VNLVQRLGYTESLMDVSKDISALNEHHLSEDFNLVQALQID-----KEKPEI 101

107 QAAAAAASQCGAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLVR 166
102 SSEGRFTLETNSSEKAGEDASTVERETQDDTNSIDSSD--DLIDGFVKI--LKEIV-- 155
167 GHESEVFICANPVSD--LLASGSDSTARINWLSNSTSGSTQ-----LVLRHCIREGG 219
156 -KLDNISVSTWNPDESILAYGEKNSVARLARIETDQEGKKYWKLIIEELRHPFALSA 214

220 QDVPSNKDVTSLDWNSEGLTLATSGYGFARINTKDGKGLASTLQGHKGP;FALKWKKGN 279
215 SSGKTTNQVTCIAWSHDCNSIVTGVENGELRLNKTGALLNVLNFRHAPIVSVKWKDGT 274

280 FILSAGVDKTTIWDATGEAKQOFPF-----HSAP---ALDWDQSNNTFAS 324
275 HIISMDVENVTILMNVI-SGTVMQHFELKETGGSGSINAENSGDGLVDVVEWDDDKFVI 334

325 CSTDMCHVCKLQGDRIKTPQGHTEVNAIKWPTGNILLASCSDDMTLKIWSMKQNCV 384
335 PGPKGALFVQITEKTTGKLIGHGPISVLEFNDTKLLLSASDDGTLLIWHGNGNSQ 394

385 HDLQHNKEIYTIKWSPTGPTNPNANMLASASFDSTVRLWVDVDRGICIHITLTKHQEP 444
395 NSFYGHSSQISVASWV-----GDDKVISSCMDGVSRLWSLKQNTLLALSIVDGV 444
445 VYSVAFSPDGRYLASGFGKCVHIW---NTQTGAL-----VHSYR 481
445 IFAGRIQDGGKYAVAFMDGVGVVYDLKLNKSRSLYGRDGLINPLPIPLVASYOSSQ 504

482 GTGGEIFVCMNAAGDKVGA--SASDGSVCVL 510
505 DNDYIFDLNWCAGNKISVAYSLOEGSVVAI 535

RESULT 6
YY46_ANASP STANDARD; PR7; 1526 AA.
AC Q8YRI1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr3466.
GN ALR3466.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Itiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RL DNA Res. 8:205-213(2001).
CC -I- SIMILARITY: Contains 16 WD repeats.
CC -I- SIMILARITY: Contains 1 pentapeptide repeat domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003593; BAB75165.1; -.
DR PIR; AC2239; AC2239.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 15.
DR PRINTS; PRO0320; GPROTEINRET.
DR PRODOM; PD000018; WD40; 8.
DR SMART; SM00320; WD40; 15.
DR PROSITE; PS00678; WD_REPEATS_1; 11.
DR PROSITE; PS50082; WD_REPEATS_2; 15.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR PROSITE; PS50837; NACHT; UNKNOWN_1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 334 376
FT DOMAIN 823 862 PENTAPEPTIDE.
FT REPEAT 862 901 WD 2.
FT REPEAT 904 945 WD 3.
FT REPEAT 945 985 WD 4.
FT REPEAT 986 1027 WD 5.
FT REPEAT 1030 1069 WD 6.
FT REPEAT 1072 1111 WD 7.
FT REPEAT 1114 1153 WD 8.
FT REPEAT 1156 1195 WD 9.
FT REPEAT 1198 1237 WD 10.
FT REPEAT 1240 1279 WD 11.
FT REPEAT 1282 1321 WD 12.
FT REPEAT 1324 1363 WD 13.
FT REPEAT 1366 1405 WD 14.
FT REPEAT 1408 1447 WD 15.
FT REPEAT 1450 1491 WD 16.
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4A738 CRC64;

Query Match 17.6%; Score 479; DB 1; Length 1526;
Best Local Similarity 29.8%; Pred. No. 1.6e-26;
Matches 117; Conservative 76; Mismatches 149; Indels 50; Gaps 10;

Qy 136 NGAHTIANNHTDMMEVDGVEIPPNKAV-VLRGHESEVFICAWNPNVSDLLASGSGDSTAR 194
Db 1127 NGV-ILANGSSDQIVRLWD--ISSKKCLYTLQHTINWNAVAFSPDGTALASGSDQTVR 1183

195 IWNLSNS-----TSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGLTLATSGYDGFAR 250
Db 1184 LWDISSKKCLYILQHTSW-----VNSVVFENPDGSLASGSDQTVR 1225

251 IW-TRDGNLSTLGOHKGP;FALKWKKGNFTLSAGVDKTTIWDATGEAKQOFPFHA 309
1226 LWEINSSKCLCTFQGHTSWNSVFNPDGSMLASGSDKTVLWDISSKCLHTFQGH- 1284

310 PALDWDQSNNTF-----ASCSTDMCHVCKLQGDRIKTPQGHTEVNAIKWDPDGN 362
1285 -----NWNSVAFNPDGSMLASGSDQTVRLWEISSKCLHTFQGHTSWNSVTFSPDGT 1339
Qy 363 LLASCSDDMTLKIWSMKQNCVHDLQHNKEIYTIKWSPTGPTNPNANMLASASFDS 422
```

DB 1340 MLASGSDQTVRLWSISSGSGCLYFLGHTNNGVIFSPDG-----AILASGSGDQ 1390
QY 423 TVRLWDVDRGICHTLTKHQEPVYAFSPDGRYLAGSPDKCVHINWNTOTGALVHSYRG 482
DB 1391 TVRLWSISSGCLYTLQGNHNNWGSIVFSPDGLASGSDQTVRLWNISGCLYTLHG 1450
QY 483 -TGGIFECVWNAAGKVCASASDSGVCVLDLR 513
DB 1451 HINSRVAFSSDGLILASGSDDETIKLWDVK 1482

RESULT 7
Y500 ANASP
ID Y500 ANASP STANDARD; PRT; 1258 AA.
AC Q8YTC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN ALR2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- SIMILARITY: Contains 15 WD repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP003590; BAB74499.1; -
DR PIR; A12155; A12155.
DR InterPro; IPR002182; NB-ARC.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00931; NB-ARC; 1.
DR Pfam; PF00400; WD40; 14.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 14.
DR PROSITE; PS00678; WD_REPEATS_1; 9.
DR PROSITE; PS00682; WD_REPEATS_2; 14.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.

FT REPEAT 55 93
FT REPEAT 640 679
FT REPEAT 640 679
FT REPEAT 682 721
FT REPEAT 724 763
FT REPEAT 766 807
FT REPEAT 809 849
FT REPEAT 850 889
FT REPEAT 892 931
FT REPEAT 934 975
FT REPEAT 976 1017
FT REPEAT 1019 1059
FT REPEAT 1060 1101
FT REPEAT 1103 1143
FT REPEAT 1144 1183

FT REPEAT 1186 1227 WD 15.
SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;
Query Match 17.0%; Score 463.5; DB 1; Length 1258;
Best Local Similarity 31.8%; Pred. No. 1.7e-25;
Matches 103; Conservative 44; Mismatches 100; Indels 77; Gaps 7;
QY 165 LRGHSEVFICAWNPSVLLASGSDSTARIWNLSGNSSTQVLVLRHCIREGQDVPS 224
DB 974 LEGHTDFYIGIAFSPDSQTLASASTDSSVRLWNI---STGQCFOILLH----- 1019
QY 225 NKDVTSLD-----NSEGTLATSGYDGFARIWT-KDGNLASTLGQKHPFALKWNKK 277
DB 1020 -----TDWYAVVPHQPQKIIATGSADCTVKLWNIISTGQCLTKLSEHSDKILGNWSPD 1073
QY 278 GNFIISAGVDKTTIWDHAITGEAKQFPHPSPALDWDQSNNTFASCTDMCIHVCKLG 337
DB 1074 GQLLASASADQSVRLWD-----CCTGRCVI----- 1099
QY 338 QDRPIKTFQHTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQOHKEIYTI 397
DB 1100 -----LRGHSNRVYSAIFSPNGEIIATCSDTQVKIWDWQGGKCLKTLTGHTNWVFDI 1152
QY 398 KWSPTGPTGNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYL 457
DB 1153 AFSPDGK-----ILASASHDQTVRIWDVNTGKCHHICIGTHLVSSVAFSPDGEV 1203
QY 458 ASGSPDKCVHINWNTOTGALVHSYR 481
DB 1204 ASGSDQTVRIWNVKTGECQLILR 1227

RESULT 8
HET1 PODAN
ID HET1 PODAN STANDARD; PRT; 1356 AA.
AC Q0808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative incompatibility protein HET-E-1.
HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009891; PubMed=7557402;
RA Saube S., Turcq B., Begueret J.;
RT "A gene responsible for vegetative incompatibility in the fungus
RT Podospora anserina encodes a protein with a GTP-binding motif and G
RT beta homologous domain.";
RL Gene 162:135-139(1995).
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,
CC HET-C.
CC -!- SIMILARITY: Contains 10 WD repeats.
CC -!- SIMILARITY: Contains 1 NACHT domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L28125; AAA85775.1; -
DR PIR; T18521; T18521.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.

```

DR PRINTS; PRO0320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD REPEATS 1; 10.
DR PROSITE; PS0082; WD REPEATS 2; 10.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW GTP-binding; Repeat; WD repeat.
FT DOMAIN 294 629 NACHT.
FT NP_BIND 300 307 GTP (POTENTIAL).
FT REPEAT 839 869 WD 1.
FT REPEAT 881 911 WD 2.
FT REPEAT 923 953 WD 3.
FT REPEAT 965 995 WD 4.
FT REPEAT 1007 1037 WD 5.
FT REPEAT 1049 1079 WD 6.
FT REPEAT 1091 1121 WD 7.
FT REPEAT 1133 1163 WD 8.
FT REPEAT 1175 1205 WD 9.
FT REPEAT 1217 1247 WD 10.
SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match 16.9%; Score 460; DB 1; Length 1356;
Best Local Similarity 27.9%; Pred. No. 3.3e-25;
Matches 124; Conservative 73; Mismatches 178; Indels 70; Gaps 12;

QY 78 IESLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQGSAGKE-----NT 130
DB 820 ISTSVVEAEWNACTQT-----LEHGSSVLSVAFADGQGVASGDDTKIKINDT 870
QY 131 ANGEENGAIHANNHTDMEYDGVVEIPPNKAVVLRGHESEVFICAWNPVSDLLASGSD 190
DB 871 ASG--TGTQT-----LEHGGSVMSVAFSPDRERVASGDD 904
QY 191 STABIWNLSENSTSGS--TQVLRHCIREGGQDVPSNKNVTSLDWNSSEGTLLATGSDGPA 249
DB 905 KTIKIW----DAASGTCQTLEHG-----GGR-----VQSVAFSPDQGVASGDDHTI 949
QY 250 RIW--TKDGNLSTLQGHKGPFFALKWKNKGNFILSAGVDKTIIDWAHTGEAKQGFPHS 308
DB 950 KIWAASGTCQTLEHGGSVLSVAFSPDQGVASGSDTKIKINDTASGTCQTLEHG 1009
QY 309 APALVDWQ--SNNTFASCTDMCHVCKLGGODRIKTFQHTNEVNAIKWDPTGNLIASC 367
DB 1010 GSVMSVAFSPDQGVASGSDTKIKINDTASGTCQTLEHGGSVMSVAFSPDQGVASG 1069
QY 368 SDDMTLKIWSMKQNCVHDLQHNKEIYTIKWSPTGPTNPNANLMLASASDSTVRLW 427
DB 1070 SDDHTIKINDAVSGTCQTLEHGGSVMSVAFSPDQ-----RVASGSDGTIKIW 1120
QY 428 DVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKCVHIWNTQTGALVHSYRGTG-I 486
DB 1121 DAASGTCQTLEHGGSVMSVAFSPDQGVASGSDTKIKINDTASGTCQTLEHGGSV 1180
QY 487 FEVCWNAAGDKVGASADGSCVCLD 511
DB 1181 QSVAFSPDQGVASGSDTKIKIND 1205

RESULT 9
YL24 ANASP
ID YL24 ANASP STANDARD; PRT; 1683 AA.
AC Q8V57;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein al12124.
GN AL12124.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: Contains 13 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003588; BAB73823.1; -
DR F1R; AF2071; AF2071.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 14.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 14.
DR PROSITE; PS00678; WD REPEATS 1; 3.
DR PROSITE; PS0082; WD REPEATS 2; 14.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109 WD 1.
FT REPEAT 1111 1150 WD 2.
FT REPEAT 1153 1192 WD 3.
FT REPEAT 1195 1236 WD 4.
FT REPEAT 1280 1318 WD 5.
FT REPEAT 1320 1359 WD 6.
FT REPEAT 1361 1400 WD 7.
FT REPEAT 1402 1441 WD 8.
FT REPEAT 1444 1483 WD 9.
FT REPEAT 1486 1525 WD 10.
FT REPEAT 1528 1567 WD 11.
FT REPEAT 1570 1609 WD 12.
FT REPEAT 1611 1650 WD 13.
SQ SEQUENCE 1683 AA; 185459 MW; F7CB361BF54F7137 CRC64;

Query Match 16.3%; Score 445.5; DB 1; Length 1683;
Best Local Similarity 28.5%; Pred. No. 4.9e-24;
Matches 109; Conservative 69; Mismatches 165; Indels 40; Gaps 6;

QY 124 AKNGENTANGENGAIHANNHTDMEYDGVVEIPPNKAVVLRGHESEVFICAWNPVSDL 183
DB 1290 SSDGKAIASASADNTIKLWNRHGLE-----TFTGSGGVYAVNFDPDSNI 1336
QY 184 LASGSDSTARIWNLSENSTSGSTQVLRHCIREGGQDVPSNKNVTSLDWNSSEGTLLATG 243
DB 1337 IASASLDNTIRLWQRPISPL-----EVLNAGSVYAVSFLHDGSIATA 1381
QY 244 SYDGFARIW--TKDGNLSTLQGHKGPFFALKWKNKGNFILSAGVDKTIIDWAHTGEAKQ 302
DB 1382 GADGNIQLWHSQDGSLLKTLPGNKA--IYGISFTPGQDLIASANADKTKWIRVDRGKALK 1440
QY 303 QPFHSAPALVDWQ--SNNTFASCTDMCHVCKLGGODRIKTFQHTNEVNAIKWDPTG 361
DB 1441 TLIGHDNEVKNVNFSPDGKTLASASDNTIKLWNRHGLE-----TFTGSGGVYAVNFDPDSNI 1500
QY 362 NLLASCSDDMTLKIWSMKQNCVHDLQHNKEIYTIKWSPTGPTNPNANLMLASASPD 421
DB 1501 KIIASASADKTIKRLWDSFSGNLKSLPAHNDLVYSVNFNPDGS-----MLASTSAD 1551
QY 422 STVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKCVHIWNTQTGALVHSYR 481
DB 1552 KTVKLWRSHDGHLLHTFSGHSNNVYSSSPDGRVIASASEDKTKWIKWIDGHLLTTLFQ 1611
QY 482 GTGGIFEVCWNAAGDKVGASASD 504

```

```

Db 1612 HQAGWMSAIFSPDGRGLISGLD 1634
RESULT 10
Y163_SYNY3
ID Y163_SYNY3 STANDARD; PRT; 1693 AA.
AC Q55563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein sil0163.
GN SL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC !- SIMILARITY: Contains 16 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D63999; BAA10064.1; -
CC PIR; S76086; S76086.
CC InterPro; IPR008941; TPR-like.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 16.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD000018; WD40; 13.
CC SMART; SM00320; WD40; 16.
CC PROSITE; PS00678; WD_REPEATS_2; 15.
CC PROSITE; PS00082; WD_REPEATS_2; 15.
CC PROSITE; PS02094; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042 WD 1.
FT REPEAT 1053 1083 WD 2.
FT REPEAT 1094 1124 WD 3.
FT REPEAT 1135 1165 WD 4.
FT REPEAT 1176 1206 WD 5.
FT REPEAT 1217 1247 WD 6.
FT REPEAT 1258 1288 WD 7.
FT REPEAT 1299 1329 WD 8.
FT REPEAT 1340 1370 WD 9.
FT REPEAT 1381 1411 WD 10.
FT REPEAT 1422 1452 WD 11.
FT REPEAT 1463 1493 WD 12.
FT REPEAT 1504 1534 WD 13.
FT REPEAT 1545 1575 WD 14.
FT REPEAT 1586 1616 WD 15.
FT REPEAT 1627 1657 WD 16.
SQ SEQUENCE 1693 AA; 0977827A0251CFF CRC64;
Query Match
Best Local Similarity 14.3%; Score 388.5; DB 1; Length 1693;
Matches 107; Conservative 52; Mismatches 138; Indels 79; Gaps 10;
QY 100 RDKLAQQQAAAAAASQGGAKNGCENTANGE--ENGHTTANNHTDMVEVDGVEI 157
DB 1362 RDKTARLTTEGECVAVLADHGWRGQGFDPGQIVTGS---ADKTAQLNVLG---- 1414

```

RESULT 11

```

HUS7_HUMAN
ID HUS7_HUMAN STANDARD; PRT; 485 AA.
AC G9NVX2; O60868; Q9BU54;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat protein HUSSY-07.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Magatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA TISSUE=Lung, and Placenta;
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]

```


RP SEQUENCE OF 400-485 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simonati B.,
RA Camata N., Zimbello R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RT yeast sequences.";
RL Yeast 18:69-80(2001).
RN [4]
RP SUBCELLULAR LOCATION.
RX MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -|- SIMILARITY: Contains 8 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK001320; BAA91621.1; -
DR EMBL; BC002884; AAH02884.2; -
DR EMBL; BC012075; AAH12075.1; -
DR EMBL; AJ005257; CAA06444.1; -
DR SWISS-2DPAGE; OSNVX2; HUMAN.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PROSITE; PS00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 7.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 325 366
FT REPEAT 370 409
FT REPEAT 412 451
FT REPEAT 454 484
SQ SEQUENCE 485 AA; 53266 MW; 0AEBE24B4957379 CRC64;
Query Match 13.4%; Score 366.5; DB 1; Length 485;
Best Local Similarity 26.8%; Pred. No. 4.7e-19;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;
QY 165 LRGHESVFIICANPVSDLLASGSGDSTARIWNLSEN----STFGSTQLVLRHCIREGGQ 220
DB 110 LEHSEAVISVAFSPTKYLASGSGDITVFWDLSTETPHTCKGRHWHVLSISWSPDGK 169
QY 221 DVPS-----NKDVTSLDW-----NSEGTLATSYDGFAR 250
DB 170 KLASGCKNGQILLWDPSTGKQVGRITAGHSKWITGLISWEPLHANPECRYVAVSSKDGSR 229
QY 251 IW-TKDGNLASTLQHGKPIFALKWKKGNFILLSAGVDKTIIDWHTGKQKQFPFH-- 307
DB 230 IWDTTAGRCRILTGHTQSTVCLRWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288
QY 308 --SAPALDLDV-----QSNNTFASCS 326
DB 289 WYNTMALSTDYALRTGAFEPAEASVNPQDLQGSQELKERALGRYLNVRQGGPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTIQGHTEVNAIKWDPGTGNLLASCSDDMTLKWSMKQDNCVH 385

DB 349 DQFTFLNSPAEDKKKPLRTMTGHOALINQVLFSPDSRIVASAFDKSIKLDGRTGKYLA 408
QY 386 DLQOENKEIYIKWSPGTGFTNNPNANMLASASPDSTVRLWDVDRGICHTLTHKHQEPV 445
DB 409 SURGHVAAVYQIANSAD-----SRLVSGSDSTLKWYDVKRAQKLANDLPQHADEV 459
QY 446 YSVAFSPDGRVSLASGDFKCVHIW 469
DB 460 YAVDMSPDGQORVASGGKDKLRIW 483
RESULT 12
HUS7 MOUSE
ID HUS7 MOUSE STANDARD; PRT; 485 AA.
AC Q8V8T4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -|- SIMILARITY: Contains 8 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BC018399; AAH18399.1; -
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 7.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282

```

FT REPEAT 197 241 WD 3.
FT REPEAT 244 282 WD 4.
FT REPEAT 327 366 WD 5.
FT REPEAT 370 409 WD 6.
FT REPEAT 412 451 WD 7.
FT REPEAT 454 484 WD 8.
SQ SEQUENCE 485 AA; 53148 MW; 207595603AEC07B CRC64;

Query Match 13.4%; Score 365.5; DB 1; Length 485;
Best Local Similarity 26.6%; Pred. No. 5.5e-19;
Matches 102; Conservative 48; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRHSESEFICAMPVPSDLASGSGDSTARINWLNSEN-----STSGSTQLVLRHCIRGGQ 220
DB 110 LEHSEAVISVAFSPTGKYLAGSGDITVRFWDLSTETPHFTCKGHRHWLSISWSPDGK 169
QY 221 DVPS-----NVDVTSLDW-----NSETLLATGSYDGFAR 250
DB 170 KLASGCKNGQILLWDPSTGLQVGRTLTGHSKWITGLSWEPLHMNPECRYVASSSKDGSVR 229
QY 251 IW-TRKGNLSTLGHCKGPIPALXWKNKGNFILSAGVDKTTIWDHTGEAKQOPPH-- 307
DB 230 VMDTTAGCERILLTGHTOSVTLRWGGDG-LLYSASQDRITKWRANDGVLCRTLOHGH 288
QY 308 --SAPALDWDV-----QSNNTFASCS 326
DB 289 WNTWTALSTDVALRTGAPEPAEATVNAQDLOGSLKELKERASSRYNLVRGGPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHNEVNAIKWDPGTGNILASCSDMTLKIWSKQDNCVH 385
DB 349 DDFTLFLWSPAEDKPLARMTGHOALINQVLFSPDSRIVASASPKSIKLDWGGTGYLA 408
QY 386 DLQCHKEIYIKWSPGTGNTNPNANILASASFDSTVLWDVDRGICHTITLTKHPEV 445
DB 409 SLRGHVAAYVQIANSAD-----SRLVSGSSDSTLKWVDKAKLATDLPGHADV 459
QY 446 YSVAFSPDGRVLASGSPKCVHIW 469
DB 460 YAVDWSPDGQVRVAGGKDKCLRIV 483

```

RESULT 13

```

YCW2_YEAST STANDARD; PRT; 515 AA.
AC P25382; Q8NKJ4;
DT 01-MAY-1992 (Rel. 22, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein YCR072C.
GN YCR072C OR YCR72C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alverghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benoit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA Botolin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Carginani G., Chanet R., Contreras R., Crouzet M., Daiguan-Fornier B.,
RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA Dujon B., Duesternost A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Franchingues-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.J., Gent M.E., Gigot D., Gillaquet V., Glansdorff N.,
RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haesemann M.,
RA Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhaus U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,

```

```

RA Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA Pereira J., Philippsen P., Pierard A., Plantia R.J., Plevani P.,
RA Poetsch B., Pohl P.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Thiries G., Triano L.N., Urrutiarazu L.A., Vallie G., Vetter I.,
RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warrington J.R.,
RA von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RA Zimmermann F.K., Sgourou J.G.;
RT "The complete DNA sequence of yeast chromosome III.";
RL Nature 357:38-46(1992).
RN [2]
RP REVISIONS.
RA Jimenez A.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISION TO 444.
RA Valles G., Volckaerts G.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59720; CAC42989.1; -.
CC PIR; S19487; S19487.
CC GERMOnline; 138973.
CC SGD; S0000668; YCR072C.
CC InterPro; IPR001632; Gpoteins_B.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 8.
CC PRINTS; PR00319; GPROTEINB.
CC PRINTS; PR00320; GPROTEINBPT.
CC ProDom; PD000018; WD40; 6.
CC SMART; SM00320; WD40; 8.
CC PROSITE; PS00578; WD_REPEATS_1; 5.
CC PROSITE; PS00582; WD_REPEATS_2; 7.
CC PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT REPEAT 141 181
FT REPEAT 184 223
FT REPEAT 227 273
FT REPEAT 276 314
FT REPEAT 352 396
FT REPEAT 400 439
FT REPEAT 442 481
FT REPEAT 484 514
FT REPEAT 515 AA; 57025 MW; 04451B0C38FB5318 CRC64;
SQ SEQUENCE 515 AA; 57025 MW; 04451B0C38FB5318 CRC64;

Query Match 13.4%; Score 365.5; DB 1; Length 515;
Best Local Similarity 23.2%; Pred. No. 6e-19;
Matches 124; Conservative 85; Mismatches 164; Indels 161; Gaps 22;

QY 38 NINGALVPPAALISITQKGLQYVEAEVSNEDGT-----FDGR-----PIESLSLIDA 86
DB 44 NYGGLRVPGA---ISEKLELLNQLNGTSDPPVTFSTCTIQGKASDPVKITIDTN 100
QY 87 VNPDVVQTRQAYRKLQAQQQAAAAAASQGSANGKNGTANGENGAIHANNHT 146
DB 101 LYSSLIKPGYNSTEDQITLLYTPRAVFKVPVTRSSA----- 138
QY 147 DMNEVDGDVIEPPNKAVALRGLHSEVFFICAWNP-VSDLLASGSGDSTARINWLNSE---N 201

```



```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=99367400; PubMed=10438340;
RA Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
RT "Structural organization, tissue expression, and chromosomal
RT localization of C10A1, a functional modulator of the Wilms' tumor
RT suppressor, WT1.";
RL Immunogenetics 49:900-905(1999).
CC -!- FUNCTION: Seems to specifically modulate the transactivation
CC activity of WT1.
CC -!- SUBUNIT: Interacts with WT1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: 'C10A' means 'bridge' in Chinese.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U63810; AAC24948.1; -;
DR EMBL; AC004020; AAC23493.1; -;
DR EMBL; BC001395; AAC01395.1; -;
DR EMBL; BC032812; AAC32812.1; -;
DR MIM; 604333; -;
DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II promoter; TAS.
DR InterPro; IPR001680; WD40.
DR Pfam; PF04000; WD40.7
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 2.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Nuclear protein.
FT REPEAT 14 53 WD 1.
FT REPEAT 59 98 WD 2.
FT REPEAT 103 142 WD 3.
FT REPEAT 148 187 WD 4.
FT REPEAT 192 231 WD 5.
FT REPEAT 250 289 WD 6.
FT REPEAT 301 338 WD 7.
SQ SEQUENCE 339 AA; 37840 MW; 63A8D8257A204FC8 CRC64;

Query Match 12.8%; Score 348; DB 1; Length 339;
Best Local Similarity 30.4%; Pred. No. 6.3e-18;
Matches 92; Conservative 42; Mismatches 89; Indels 80; Gaps 11;

```

```

QY 222 VPSNKDVTS--LDWNSGTLATGYDGFARITWKDGN---LASTLGO-HKGPIFALKWN 275
DB 11 VFAHPDSRCWFLAWNPAAGTLLASCGDRIRIWTGTEGDSWICKSVLSRSHQRTVRKVAWS 70
QY 276 KKGNFILSAGVDKTTIIWDAHTGEAKQOPPHSAAPALDQVDSNNTPASCSTDMCIHVCK 335
DB 71 PCGNVLASASFDATTCI-----WKQDQDFECVT----- 99
QY 336 LQQDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKINSMKQDN---CVHDLQOHNK 392
DB 100 -----TLECHENEKSVAVAPSGNLLATCSRDKSVWVWEVDEDEYEYECVSLNSHTQ 151
QY 393 EYITIKWSTGPTGNPNANMLASASFDSTVRLW--DVDRGICHTLTKHQEPVYSVAF 450
DB 152 DVKHVVVWHP-----SQELLASASYDDTVKLYREBEDDWVCCATLEGHSTVWSLAF 202
QY 451 SPDGRYLASGSPDKCKVHIWNTQTGALVHSYRGTTGGIFEVCVWNAAGDKYVGASASDGS---V 507
DB 203 DFGQRLASCSDDRTVRIWRYLPG-----NEQG--VACSGSDPSWKCI 244
QY 508 CVL 510
DB 245 CTL 247

Search completed: August 9, 2004, 16:46:10
Job time : 18.6667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 2731

Sequence: 1 MSISDEVNFIYRLQSSG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2696	98.7	514	TELR_HUMAN	Q9bzk7 homo sapien
2	2439	89.3	526	TEBLX_HUMAN	O60907 homo sapien
3	2351	86.1	522	TEBLX_HUMAN	Q9bq87 homo sapien
4	1304	47.7	313	TEBLX_MOUSE	Q9qxe7 mus musculus
5	501.5	18.4	535	SIF2_YEAST	P38262 saccharomyc
6	481	17.6	1526	VY46_ANASP	Q8yrr1 anabaena sp
7	462	16.9	1356	HET1_PODAN	Q08088 podospora a
8	439.5	16.8	1258	Y500_ANASP	Q8ytc2 anabaena sp
9	451	16.5	1683	YL24_ANASP	Q8yvt7 anabaena sp
10	390.5	14.3	1693	YL163_SYNY3	Q55563 synechocyst
11	375.5	13.7	515	YCW2_YEAST	P25382 saccharomyc
12	363.5	13.3	485	HUS7_HUMAN	Q9nvx2 homo sapien
13	362.5	13.3	485	HUS7_MOUSE	Q8vej4 mus musculus
14	358	13.1	437	ACAC3_DICDI	P4197 dictyosteli
15	349	12.8	339	CIAL1_HUMAN	O76071 homo sapien
16	338.5	12.4	742	PKWA_THRCU	P49695 thermomonos
17	336.5	12.3	409	LIS1_HUMAN	P43034 homo sapien
18	335.5	12.3	409	LIS1_BOVIN	P43033 bos taurus
19	335.5	12.3	800	T2D4_HUMAN	Q15542 homo sapien
20	334.5	12.2	409	LIS1_MOUSE	P43035 mus musculus
21	331.5	12.1	334	WDR5_HUMAN	Q9g9p9 homo sapien
22	328.5	12.0	361	WDR5_DROME	Q9v378 drosophila
23	328.5	12.0	606	PF20_CHLRE	P93107 chlamydomon
24	326	11.9	798	T2D4_YEAST	P38129 saccharomyc
25	324.5	11.9	579	SE10_CABEL	Q93794 caenorhabdi
26	313	11.5	640	MT30_YEAST	P39014 saccharomyc
27	312	11.4	376	YKY4_CABEL	P39014 caenorhabdi
28	310	11.4	1191	YL443_SYNY3	Q17963 synechocyst
29	307	11.2	643	T2D4_SCHPO	O13282 schizosacch
30	306.5	11.2	704	T2D4_DROME	P49846 drosophila
31	306	11.2	678	SCOB_EMENI	Q00659 emericaella
32	303.5	11.1	501	YH92_CABEL	Q23256 caenorhabdi
33	302	11.1	605	POFI1_SCHPO	P87053 schizosacch

ALIGNMENTS

RESULT 1

ID	TELR_HUMAN	STANDARD;	PRT;	514 AA.
AC	Q9BZK7; Q9H2I1; Q9H9A1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBL1-related protein 1).			
GN	TBLR1 OR IRA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.			
RX	MEDLINE=21930350; PubMed=11931768;			
RA	Zhang J., Kalkum M., Chait B.T., Roeder R.G.;			
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GFS2.";			
RL	Mol. Cell 9:611-623 (2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Zhang X., Dormady S., Basch R.;			
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,			
RA	Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,			
RA	Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;			
RT	"NEDO human cDNA sequencing project.";			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
CC	- - SUBUNIT: Interacts with HDAC3.			
CC	- - SIMILARITY: Contains 1 Lish domain.			
CC	- - SIMILARITY: Contains 8 WD repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF314544; AAK00301.1; -			
DR	EMBL; AF268193; AAG44736.1; -			
DR	EMBL; AK022956; BAB14331.1; -			
DR	InterPro; IPR006594; LISH.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 8.			
DR	PRINTS; PR00320; GPROTEINBRPT.			
DR	ProDom; PD000018; WD40; 3.			
DR	SMART; SM00667; LISH; 1.			

Q09990 caenorhabdi
O14775 homo sapien
P54314 mus musculu
Q93847 caenorhabdi
P49178 zea mays (m
Q09715 schizosacch
Q9uug8 schizosacch
P04901 homo sapien
Q9fuy2 arabidopsis
P54311 rattus norv
Q40687 oryza sativ
O22212 arabidopsis

1 LI23_CABEL
1 GBB5_HUMAN
1 GBB5_MOUSE
1 YZLL_CABEL
395 10.7
380 10.6
614 10.5
586 10.5
340 10.5
931 10.4
340 10.4
380 10.4
554 10.4

34 299
35 294.5
36 292.5
37 292
38 291
39 288.5
40 288
41 295.5
42 285
43 284.5
44 283
45 283

SMART: SM00320; WD40; 8.
PROSITE; PS00696; LISH; 1.
PROSITE; PS00678; WD_REPEATS_1; 4.
PROSITE; PS00682; WD_REPEATS_2; 6.
PROSITE; PS0294; WD_REPEATS_REGION; 1.
Repeat; WD repeat.
DOMAIN 4
FT REPEAT 36 LISH.
FT REPEAT 167 206 WD 1.
FT REPEAT 223 262 WD 2.
FT REPEAT 264 303 WD 3.
FT REPEAT 306 344 WD 4.
FT REPEAT 347 386 WD 5.
FT REPEAT 389 437 WD 6.
FT REPEAT 440 479 WD 7.
FT REPEAT 481 513 WD 8.
DOMAIN 108 118 POLY-ALA.
FT CONFLICT 31 31 E -> K (IN REF. 2).
FT CONFLICT 59 59 Y -> H (IN REF. 3).
FT CONFLICT 389 389 A -> Q (IN REF. 2).
SEQUENCE 514 AA; 55559 MW; 0B556D2E4BA796D CRC64;
Query Match 98.7%; Score 2696; DB 1; Length 514;
Best Local Similarity 98.8%; Pred. No. 5.7e-188;
Matches 508; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSEYVNFVLYRQESGFSHSAFTGIESHSQSNGINGALVPPAALISIIQKGLQV 60
DB 1 MSISSEYVNFVLYRQESGFSHSAFTGIESHSQSNGINGALVPPAALISIIQKGLQV 60
QY 61 EAEVSEINDEGTLFDGRPTESLSLIDAVPDVQVQQAQRDKLAQQAQAAAAAAATNQ 120
DB 61 EAEVSEINDEGTLFDGRPTESLSLIDAVPDVQVQQAQRDKLAQQAQAAAAAAASQ 120
QY 121 QQSAXNGENTANGENGATIANNHDTMMVEVDGVEIPSNKAVVLRGHESEVFIQAWNPV 180
DB 121 QQSAXNGENTANGENGATIANNHDTMMVEVDGVEIPPNKAVVLRGHESEVFIQAWNPV 180
QY 181 SLLVSGSDSTARIWNLSENSTGPTOLVLRHCIRGGQDVPSNKDVTSLDWNSGTL 240
DB 181 SLLVSGSDSTARIWNLSENSTGPTOLVLRHCIRGGQDVPSNKDVTSLDWNSGTL 240
QY 241 ATGSDYDGPARIWTKDGNLSTLQGHKGFIFALKWNKNGNFTLSAGVDKTTIINDAHTGEA 300
DB 241 ATGSDYDGPARIWTKDGNLSTLQGHKGFIFALKWNKNGNFTLSAGVDKTTIINDAHTGEA 300
QY 301 KQOFFPHSAPALVDQWQNNFTFASCTDMCTHVCKLQGDRIKPTFGQHTNEVNAIKWDPT 360
DB 301 KQOFFPHSAPALVDQWQNNFTFASCTDMCTHVCKLQGDRIKPTFGQHTNEVNAIKWDPT 360
QY 361 GNLLASCDSDMTLTKWSKQDNCVHDLQAHKEIYTIKWSPTGPTNPNANMLASASF 420
DB 361 GNLLASCDSDMTLTKWSKQDNCVHDLQAHKEIYTIKWSPTGPTNPNANMLASASF 420
QY 421 DSTVRLWDVDRGICHTLTKHQEPYVVAFPDGRYLASGDFKCVHINWQTGALVHSY 480
DB 421 DSTVRLWDVDRGICHTLTKHQEPYVVAFPDGRYLASGDFKCVHINWQTGALVHSY 480
QY 481 RGTGGIFVYCNAAAGDKVGCASASGSCVLDLRK 514
DB 481 RGTGGIFVYCNAAAGDKVGCASASGSCVLDLRK 514

RESULT 2
TBLX_HUMAN STANDARD; PRT; 526 AA.
ID TBLX_HUMAN AC O60907;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transducin beta-like 1X protein (transducin-beta-like 1, x-linked).
GN TBLX OR TBL1.
OS Homo sapiens (Human).

```
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 POLY-ALA.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match
  89.3%; Score 2439; DB 1; Length 526;
Best Local Similarity 86.0%; Pred.No.2.5e-169;
Matches 453; Conservative 33; Mismatches 21; Indels 14; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSHSAFTGIESHISQSNGALVPPAALISIIQKGLQYV 60
DB 1 MSITSEVNLFVRYLQESGFSHSAFTGIESHISQSNGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSDNEGTGILFDCGRPEISLIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 118
DB 61 EAEISINEDGVDFDRPEISLIDAVMPDVVQTRQOAFREKLAQOHAHAHAHAATAA 120
QY 119 -----NQOQSAKNGENTANGEENGAAHTIANHNTDMVEVDGVEIIPSKAVLRLG 167
DB 121 ATAATTSGAGVSHQNPKNREATVNGEENRAHSV-NNHAKPVEIDGVEIIPSKATVLRG 179
QY 168 HESEVFIICAWNPVSDLLVSGSDSTARIWNLSENSTSGPTOLVLRHCIREGGQDVPSNKD 227
DB 180 HESEVFIICAWNPVSDLLVSGSDSTARIWNLSENSTSGPTOLVLRHCIREGGQDVPSNKD 239
QY 228 VTSIDWNSSEGTLATGSDGFARITWKGDLASTLGQHKGFIFALKWKNKGNFISLQAGVD 287
DB 240 VTSIDWNTGTLATGSDGFARITWKGDLASTLGQHKGFIFALKWKNKGNFISLQAGVD 299
QY 288 KTTIWDHAHTGEAKQOFPFHSAAPALVDVQNNNTFASCTDMCHIVCKLGDRIPIKTFQG 347
DB 300 KTTIWDHAHTGEAKQOFPFHSAAPALVDVQNNNTFASCTDMCHIVCKLGDRIPIKTFQG 359
QY 348 HTNEVNAIKWPTGNTGLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGNT 407
DB 360 HTNEVNAIKWPTGNTGLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGNT 419
QY 408 NPANMLASASFDSTVRLWLDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 467
DB 420 NPNSNIMLASASFDSTVRLWLDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 479
QY 468 IWNTOGTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDLRK 514
DB 480 IWNTOGTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDLRK 526

RESULT 3
ID TELX_HUMAN
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBL1Y OR TBL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC -----

or send an email to license@isb-sib.ch.
-----
CC EMBL; AF332220; AAK13472.1; -
CC EMBL; AF332221; AAK13473.1; -
CC EMBL; AF332222; AAK13474.1; -
CC Genew; HGNC:18502; TBL1Y.
CC InterPro; IPR006594; Lish.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 8.
CC PRINTS; PR00320; GPROTEINERPT.
CC ProDom; PD000018; WD40; 3.
CC SMART; SM00667; Lish; 1.
CC SMART; SM00320; WD40; 8.
CC PROSITE; PS00896; Lish; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS50082; WD_REPEATS_2; 6.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match
  86.1%; Score 2351; DB 1; Length 522;
Best Local Similarity 83.1%; Pred.No.5.9e-163;
Matches 439; Conservative 35; Mismatches 37; Indels 12; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSHSAFTGIESHISQSNGALVPPAALISIIQKGLQYV 60
DB 1 MSITSEVNLFVRYLQESGFSHSAFTGIESHISQSNGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSDNEGTGILFDCGRPEISLIDAVMPDVVQTRQOAYRDKLAQOHAHAHAHAAT-- 115
DB 61 EAEISINEDGVDFDRPEISLIDAVMPDVVQTRQOAFREKLAQOHAHAHAHAATAA 120
QY 116 -----AATNQOQSAKNGENTANGEENGAAHTIANHNTDMVEVDGVEIIPSKAVLRLGHE 169
DB 121 AATMTPAALISQQNPKNREATVNGEENGAHEI-NNHAKPVEIDGVEIIPSKATVLRGHE 179
QY 170 SEVFIICAWNPVSDLLVSGSDSTARIWNLSENSTSGPTOLVLRHCIREGGQDVPSNKDVT 229
DB 180 SEVFIICAWNPVSDLLVSGSDSTARIWNLSENSTSGPTOLVLRHCIREGGQDVPSNKDVT 239
QY 230 SLDWNSSEGTLATGSDGFARITWKGDLASTLGQHKGFIFALKWKNKGNFISLQAGVDKT 289
DB 240 SLDWNSDGTLTLAGSYDGFARITWENGLASTLGQHKGFIFALKWKNKGNFISLQAGVDKT 299
QY 290 TTIWDHAHTGEAKQOFPFHSAAPALVDVQNNNTFASCTDMCHIVCKLGDRIPIKTFQHT 349
DB 300 TTIWDHAHTGEAKQOFPFHSAAPALVDVQNNNTFASCTDMCHIVCKLGDRIPIKTFQHT 359
QY 350 NEVNAIKWPTGNTGLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGNTNP 409
DB 360 NEVNAIKWPTGNTGLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGNTNP 419
QY 410 NANMLASASFDSTVRLWLDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIW 469
DB 420 NSSIMLASASFDSTVRLWLDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIW 479
QY 470 NTCGTGALVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDL 512
DB 480 NTCGSGLVHSYRGTFGIFVCNNAAGDKVGASASDGSVCVLDL 522

RESULT 4
ID TELX_MOUSE
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transducin beta-like 1Y protein (Transducin-beta-like 1, Y-linked).
GN TBL1Y OR TBL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC -----
```



```
DR PROSITE; PS50082; WD_REPEATS_2; 4.  
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.  
KW Repeat; WD repeat; Chromatin regulator; Nuclear protein.  
FT DOMAIN 4 36 LISH.  
FT REPEAT 163 208 WD 1.  
FT REPEAT 218 257 WD 2.  
FT REPEAT 259 298 WD 3.  
FT REPEAT 315 354 WD 4.  
FT REPEAT 357 396 WD 5.  
FT REPEAT 399 437 WD 6.  
FT REPEAT 440 479 WD 7.  
FT CONFLICT 396 396 S -> C (IN REF. 1; CA85058).  
SQ SEQUENCE 535 AA; 59145 MW; 25DD19AB29FB4B07 CRC64;  
  
Query Match 18.4%; Score 501.5; DB 1; Length 535;  
Best Local Similarity 25.4%; Pred. No. 6.8e-29;  
Matches 143; Conservative 101; Mismatches 23; Indels 79; Gaps 16;  
  
QY 1 MSISDEVNFTVRYLQSGFSGHAFTEGIESHSQSIN-GALVPPAALSIIOKQY 59  
DB 1 MSITSEENLYIRYCQENGHEVSALAQDTRVLEFDEKYEHIPLGTLVNLVQRGILY 60  
  
QY 60 VEAEVSNEDG---TLFDGRPIESLSLIDAVMPVQTRQAYRDKLAQQHAAAAA 116  
DB 61 TESELMVDSKGDIGALSNEHLSDFNLVQALQID-----KEKPE--ISSEGRFTL 109  
  
QY 117 ATNQGSAKNGENTANGENGAAHTIANNHTDM-EVGDVEIPSNKAVILRGHSEVPIC 175  
DB 110 ETNSE-SNKADEDGASTVERTEQDDTSIDSSDLDFGFKI--LKSIV---KLDNIYSS 163  
  
QY 176 AWPVSD-LLVSGSGDSTARIWLSNSENSTGPTQ-----LVLRHCIREGQDVPSNKDV 228  
DB 164 TWPPLDESILAYGEKNSVARLARIVETDQEGKYWKLTIIAELRHPPALSASSGKTTNQV 223  
  
QY 229 TSLDWNSEGLTAGSDGDFARIWTKDGNLASTLGQHKGPFAKWNKGNFILSAGVDK 288  
DB 224 TCLAWSHDGNSIVTGVENGELELWKTGALLNLNFRHAPTIVSVKWKNGDTHIIMSDVEN 283  
  
QY 289 TTIWDAHTGAQKQFPF-----HSAP-----ALDVMQSNNTFASCSDDMCIVH 333  
DB 284 VTILNNVISIVMGHFLKETGGSSINAEHSGSLGVDVWDDKFIPIPGKGAIFV 343  
  
QY 334 CKLGQDRPIKTFQGHTEVNAIKWDPGNLILASCDMTLKIKWMDKNCVHDLQAHNKE 393  
DB 344 YQITEKTPTKLIGHGPIISVLEFNDINKLLSASDDGTIRIWHGGNGNSFYGHQS 403  
  
QY 394 IYTIKWSPTGTTNPNANMLASFDSTVRLWDVRGICIHLLTKHQEPVYSVAFSPD 453  
DB 404 IVSASWV-----GDDKVIKSCMDGVSRLMSLKONTLLALSIVDGVPIFAGRISQD 453  
  
QY 454 GRYLASGDFDKCVHIW-----NTQTGAL-----VHSYRGTGIGIFEVC 490  
DB 454 GQYAVAFMDQGVNVYDLKLNSKRSRLYGNRDGILNPLPIPLVYASQSSQDNDYFDLS 513  
  
QY 491 WNAAGDKVGA--SASDGSVCVL 510  
DB 514 WNCAGNKISVAYSLOQGSVVAI 535  
  
RESULT 6  
YY46 ANASP  
ID YY46 ANASP  
AC Q8YRII; STANDARD; PRT; 1526 AA.  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical WD-repeat protein alr3466.  
GN ALR3466.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Katanabe A., Iritiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
CC -I- SIMILARITY: Contains 16 WD repeats.  
CC -I- SIMILARITY: Contains 1 pentapeptide repeat domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AP003593; BAB75165.1; -;  
DR PIR; AC2239; AC2239.  
DR InterPro; IPR007111; NACHT_NTPase.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 15.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR PRODOM; PD000018; WD40; 8.  
DR SMART; SMC00320; WD40; 15.  
DR PROSITE; PS00678; WD_REPEATS_1; 11.  
DR PROSITE; PS50082; WD_REPEATS_2; 15.  
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.  
DR PROSITE; PS50837; NACHT; UNKNOWN 1.  
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.  
FT REPEAT 334 376  
FT DOMAIN 823 862 PENTAPEPTIDE.  
FT REPEAT 862 901 WD 2.  
FT REPEAT 904 945 WD 3.  
FT REPEAT 946 985 WD 4.  
FT REPEAT 988 1027 WD 5.  
FT REPEAT 1030 1069 WD 6.  
FT REPEAT 1072 1111 WD 7.  
FT REPEAT 1114 1153 WD 8.  
FT REPEAT 1156 1195 WD 9.  
FT REPEAT 1198 1237 WD 10.  
FT REPEAT 1240 1279 WD 11.  
FT REPEAT 1282 1321 WD 12.  
FT REPEAT 1324 1363 WD 13.  
FT REPEAT 1366 1405 WD 14.  
FT REPEAT 1408 1447 WD 15.  
FT REPEAT 1450 1491 WD 16.  
SQ SEQUENCE 1526 AA; 170376 MW; D5DF272509B4A738 CRC64;  
  
Query Match 17.6%; Score 481; DB 1; Length 1526;  
Best Local Similarity 30.3%; Pred. No. 8.6e-27;  
Matches 118; Conservative 77; Mismatches 149; Indels 46; Gaps 11;  
  
QY 136 NGAHTIANNHTDMVEVDGVEIPSNKAV-VLRGHESEVFICAWNPNVSDLLVSGSGDSTAR 194  
DB 1127 NGV-TLNGSSDQIVRLWD--ISSKKCLYTLQHTNWNNAVAFPDGATLASGSDQTVR 1183  
  
QY 195 IWNLSNENSTGPTQLVLRHC--IREGGQDVPSNKDVTSLDWNSEGLTAGSYDGFARIW 252  
DB 1184 LWDISSK-----CLYLQG-----HTSVNVSVFNPDGSLASGSDQTVRLW 1227  
  
QY 253 -TKDGNLASTLGQHKGPFAKWNKGNFILSAGVDKTTIWDATGAKQFPFHSAPA 311  
DB 1228 EINSSKCLCTFGHTSWNSVFNPDGSLASGSDKTVRLWDISSKCLCTFGQHT--- 1284  
  
QY 312 LDVDMQSNNTF-----ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTLNLL 364  
DB 1285 ---NWVNSVAFNPDGSLASGSDQTVRLWEISSKCLCTFGHTSWVSVTFSPDGTML 1341  
  
QY 365 ASCSDMTLKIKWMDKNCVHDLQAHNKEIYTIKWSPTGTTNPNANMLASAFDSTV 424
```



```

Db 1612 HQAGVMSAIFSPDGKTLISGLD 1634
RESULT 10
Y163 SYN3
ID Y163 SYN3 STANDARD; PRT; 1693 AA.
AC Q55563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein sil0163.
GN SLI0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_taxid=1148;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- SIMILARITY: Contains 16 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D63999; BAA10064.1; -.
DR PIR; S76086; S76086.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40.16.
DR PRINTS; PR00320; GPOTENBRPT.
DR ProDom; PD000018; WD40.13.
DR SMART; SM00320; WD40.16.
DR PROSITE; PS00678; WD REPEATS_1; 8.
DR PROSITE; PS0082; WD REPEATS_2; 15.
DR PROSITE; PS0294; WD REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042
FT REPEAT 1053 1083
FT REPEAT 1094 1124
FT REPEAT 1135 1165
FT REPEAT 1176 1206
FT REPEAT 1217 1247
FT REPEAT 1258 1288
FT REPEAT 1299 1329
FT REPEAT 1340 1370
FT REPEAT 1381 1411
FT REPEAT 1422 1452
FT REPEAT 1463 1493
FT REPEAT 1504 1534
FT REPEAT 1545 1575
FT REPEAT 1586 1616
FT REPEAT 1627 1657
SQ SEQUENCE 1693 AA; 189935 MW; 0977A827A0251CFF CRC64;
Query Match 14.3%; Score 390.5; DB 1; Length 1693;
Best Local Similarity 28.5%; Pred. No. 3.6e-20;
Matches 107; Conservative 53; Mismatches 137; Indels 79; Gaps 10;
QY 100 RDKLAQQHAAHAAAAATNOGSAKNGENTANGE--ENGAAHTIANNHDTMWEYDGDVEI 157
DB 1362 RDKTARLTTEGECVAVLADHQGWREGQFSPDGQWIVTGS---ADKTAQLMNVLG---- 1414

```

RESULT 11

```

YCW2 YEAST
ID YCW2 YEAST STANDARD; PRT; 515 AA.
AC P25382; Q8NJK4;
DT 01-MAY-1992 (Rel. 22, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical WD-repeat protein YCR072C.
GN YCR072C OR YCR72C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Francinques-Gaillard M.-C., Franco L., Prontali L., Fukuhara H.,
RA Fuller L.J., Gent M.E., Gigot D., Gillet V., Glansdorff N.,
RA Goffeau A., Genson M., Grisanti P., Grivell L.A., Haasemann M.,
RA Hata D., Hegemann J.H., Herbert C.J., Hilger F., Hohnann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhaus U., Kreisl P., Lafranchi G., Lewis C., van der Linden C.G.,
RA Lucchini G., Lutzenkirchen K., Maat C., Manhaupt G., Manzano M.E.,
RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
RA Perea J., Philippsen P., Paller A., Planter R.J., Plevani P.,
RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA Sanz X., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Thirios G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warming J.R.,
RA von Wettstein D., Wickstead B.L., Wilson C., Wurst H., Xu G.,
RA Zimmermann P.K., Sgouros J.G.;
RT "The complete DNA sequence of yeast chromosome III.";
RL Nature 357:38-46(1992).

```



```

RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001320; BAA91621.1; -
DR EMBL; BC002884; AAH02884.2; -
DR EMBL; BC012075; AAH12075.1; -
DR EMBL; AJ005257; CAA06444.1; -
DR SWISS-2DPAGE; Q9NVX2; HUMAN.
DR InterPro; IPR001632; Gpotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00682; WD REPEATS 2; 7.
DR PROSITE; PS50082; WD REPEATS 3; 7.
DR PROSITE; PS50294; WD REPEATS_REGION; 1.
KW Nuclear protein; Repeat; WD Repeat.
FT REPEAT 112 151
FT REPEAT 151 193
FT REPEAT 193 241
FT REPEAT 241 282
FT REPEAT 282 325
FT REPEAT 325 366
FT REPEAT 366 409
FT REPEAT 409 451
FT REPEAT 451 484
FT REPEAT 484 532
SQ SEQUENCE 485 AA; 53266 MW; 0AEBE24B449573799 CRC64;
Query Match 13.3%; Score 363.5; DB 1; Length 485;
Best Local Similarity 26.3%; Pred. No. 6e-19;
Matches 102; Conservative 47; Mismatches 146; Indels 89; Gaps 9;
QY 165 LRGHSEVFICAWNPVSDLLVSGSGDSTARIWNLSEN-----STSGPTQLVLRHCIRGGQ 220
DB 110 LEGHSEAVISAFSPSTGKYLASGSGDTTFRFWDLSLTETPFTCKGHRHWLISLSWSPDGK 169
QY 221 DVPS-----NKGDTSLDW-----NSEGTLATGSDYDGFAR 250
DB 170 KLASCKKQGLLDWPSGKQVGRTRLAGSKWITGLSWEPLHNPCEYVASSSKDGSVR 229
QY 251 IW-TKDGNLASTLGQHGKPIFALKWKNKGNFILSGVDKVTIWDATGGEAKQFPFH-- 307
DB 230 IWDTTAGRCERILTGTQSVTLRWGGDG-LLYSASQDRTIKVRAHDGVLCTRLQGHG 288
QY 308 --SAPALVDW-----QSNNTFASCS 326
DB 289 WYNTWALSTDAIRTAEPAEASVNPDOLOGSLQELKERALSRYNLVGGGPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQHTNEVNAIKWDPGTGNLLASCSDMTLKIWSMKQDNCVH 385
DB 349 DDFTLFLSPAEKDKKPLRTMTGHOALINQVLFSPDSRIVASASFDKSIKLDGRTGKYLA 408
QY 386 DLQANKKIYITKWSPTGPTGNPNANLKLASAFSDSTVRLWDYDRGICIIHTLTKHOEPV 445
DB 409 SLRGHVAAYQIAWSAD-----SRLLVSGSGDSTLTKWDVKAQKLAWDLPGHADEV 459
QY 446 YSAFSPDGRYLASGSKFDCVHIW 469
DB 460 YAVDASPDQRVASGSKDKCLRIW 483

```

RESULT 13
HUS7_MOUSE

```

ID HUS7_MOUSE STANDARD; PRT; 485 AA.
AC Q8VEI4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC018399; AAH18399.1; -
DR InterPro; IPR001632; Gpotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00682; WD REPEATS 2; 7.
DR PROSITE; PS50082; WD REPEATS_REGION; 1.
DR PROSITE; PS50294; WD REPEATS_REGION; 1.
KW Nuclear protein; Repeat; WD Repeat.
FT REPEAT 112 151
FT REPEAT 151 193
FT REPEAT 193 241
FT REPEAT 241 282
FT REPEAT 282 325
FT REPEAT 325 366
FT REPEAT 366 409
FT REPEAT 409 451
FT REPEAT 451 484
FT REPEAT 484 532
SQ SEQUENCE 485 AA; 53148 MW; 2D7F59D603AEC07B CRC64;
Query Match 13.3%; Score 362.5; DB 1; Length 485;
Best Local Similarity 26.3%; Pred. No. 7.1e-19;
Matches 101; Conservative 48; Mismatches 146; Indels 89; Gaps 9;
QY 165 LRGHSEVFICAWNPVSDLLVSGSGDSTARIWNLSEN-----STSGPTQLVLRHCIRGGQ 220

```

Db 110 LEGHSEAVISVAPSPGKYLASGSDTTRVFDLSTETPHCTCKGHRHWLISWSPDGK 169
Qy 221 DVPS-----NKDVTSLDW-----NSETLLATSGYDGFAR 250
Db 170 KLASGCKNGQILLWDPSTGLQVGRITLGHKSWITGLSWEPLHNPCECRYVASSSKDGSVR 229
Qy 251 IW-TKDGNLASTLGHKGPPIFALKWKNKGNFILSAGVDKTTIWDATHTGEAKQFPFH-- 307
Db 230 VWDITAGRCERILTGHSTVTCLERWGGDG-LLISASQDRITIKVRAHDGVLCRTLQGHG 288
Qy 308 --SAPALDVM-----QSNNTFASCS 326
Db 289 WNTMTALSTDYALRTGAFPAEATVNAQDLQGLSKELKERASSRYNLVRGQPERLVSGS 348
Qy 327 TDMCIHVKCLQD-RPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTLKIKWSMKQDNCVH 385
Db 349 DDFTLFWSPADKPKPLARMTGHQALNQVLFSPDSRIVASASPKSIKLWDGRTGKYLA 408
Qy 386 DLQAHNKEIYTIKSPGPGTNNPNANMLASAFSDTVLWDRGICHTITLKHOEPV 445
Db 409 SLRGHVAAYQIAWSAD-----SRLLVSGSDSTLKVWQVKAQKLATDLPGHADV 459
Qy 446 YSVAFSPDGRYLASGSPDKCVHIW'459
Db 460 YAVDWSPDQGVASGGKDKLRIW 483

RESULT 14

AAC3_DICDI
ID -AAC3_DICDI STANDARD; PRT; 437 AA.
AC P14197;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE AAC-rich mRNA clone AAC3 protein (Fragment).
GN AAC3.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90066348; PubMed=2511421;
RA Shaw D.R., Richter H., Giorda R., Ohmachi T., Ennis H.L.;
RT "Nucleotide sequences of Dictyostelium discoideum developmentally
RT regulated cDNAs rich in (AAC) imply proteins that contain clusters of
RT asparagine, glutamine, or threonine."
RL Mol. Gen. Genet. 218:453-459 (1989).
CC -!- DEVELOPMENTAL STAGE: The concentration of AAC-rich mRNAs is low
CC in dormant spores and growing cells, but increases during
CC spore-germination and multicellular development.
CC -!- MISCELLANEOUS: Several proteins derive from AAC-rich mRNA, which,
CC due to a frameshift also have ACA and CAA codons and thus are
CC Asn-, Thr- or Gln-rich.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; X16524; CAA34531.1; -.
DR PIR; S05357; S05357.
DR DictyBase; DDB0001891; AAC3.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR PRINTS; PR00320; GPROTINBRPT.
DR SMART; SMC0320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.
FT NON_TER 1
FT REPEAT 122 161 WD 1.
FT REPEAT 185 227 WD 2.
FT REPEAT 229 266 WD 3.
FT REPEAT 269 308 WD 4.
FT REPEAT 316 355 WD 5.
FT REPEAT 358 397 WD 6.
FT REPEAT 399 436 WD 7.
FT DOMAIN 8 75 GLN-RICH.
SQ SEQUENCE 437 AA; 49151 MW; 755AE61228D0A1A9 CRC64;

Query Match 13.1%; Score 358; DB 1; Length 437;
Best Local Similarity 23.7%; Pred. No. 1.3e-18;
Matches 109; Conservative 79; Mismatches 175; Indels 96; Gaps 13;

Qy 93 QTRQAYRDKLAQQHAAAAAATNQGSAKNGENTANGEEGAHTIANNHTDMEVD 152
Db 28 QTQVOQLHNLHQHNNQIQQAQATQCHLOQTQYLSQIHQSQSQSLNN----- 79
Qy 153 GDVEIPSNKAVVLRGHESEVFICANVPYSDLLVSGSGDSTARIWNLSENST-----SGP 206
Db 80 -----LNSNKEST-----NIPKINTOYTNFDEKN 104
Qy 207 TQVLRLHCIREGQD-VPSNKDVTSLDMNSECTLLATGSDGFARIW----- 252
Db 105 LDLASRYFSECTKDFIGNKKKSTSVANVANGTKIASSGSDGIVRVWNPDP LGSNNNNN 164
Qy 253 -----TKDGNLST--LGQHKGP FALKWKNKGNFIL-SAGVDKTTIWDATHTGEAKQ 302
Db 165 SNTTSNSKNNNIKETIELKGDHGSIEKISWSPKNNDDLASAGTDKVIKWVDKIGKCG 224
Qy 303 QPFFHSAPALDWDQSNNTF-ASCSTDMCIHVCKLQDRPIKTFQGHTEVNAIKWDPTG 361
Db 225 TVSTNS-ENIDVRWSPDQGFIVACTRDDHALIDLPITKTLAIYKFNGBELNQVGDWNG 283
Qy 362 NLLASCSDDMTLKWSMKQDNCVH-----DLQAHNKEIYTIKSPGPGTNNPNANMLA 416
Db 284 DLILMANSNGNIEAYKFLPKSTTHVKHLTKLYGHTASICYMEFDETKG-----YLA 334
Qy 417 SASFDSVRLWVDVDRGICHTITLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTAL 476
Db 335 AGSASISVSLWIEDMMCVKTIKSTFFCRSVSFSFDGQFIASSFESTIEFIHIESOP 394
Qy 477 VHSYRGTGIFVWCN-----AAGDKV GASASDGSVCV 509
Db 395 IHTIE-CISGVSSLMWHPTLP LLAYAPEINENNKDPSIRV 432

RESULT 15

CIAL_HUMAN
ID CIAL_HUMAN STANDARD; PRT; 339 AA.
AC O76071;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat containing protein Cial 1.
GN CIAO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=98225157; PubMed=9556563;
RA Johnstone R.W., Wang J., Tommerup N., Vissing H., Roberts T., Shi Y.;
RT "Ciao 1 is a novel WD40 protein that interacts with the tumor
RT suppressor protein wt1."
RL J. Biol. Chem. 273:10880-10887 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Adams M.D., Loftus B.J., Zhou L., Crosby M., Fuhrmann J., Mason T.M.,
RA Brandon R., Kim U.J., Kerlavage A.R., Venter J.C.;

Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 (3)
 RN SEQUENCE FROM N.A.
 RP TISSUE=Blood, and Skin;
 RC MEDLINE=22384257; PubMed=12477932;
 RA Klausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Stachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 FT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=99367400; PubMed=10438340;
 RA Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
 RT "Structural organization, tissue expression, and chromosomal
 localization of Ciao 1, a functional modulator of the Wilms' tumor
 suppressor, Wt1."
 RT Immunogenetics 49:900-905 (1999).
 CC -!- FUNCTION: Seems to specifically modulate the transactivation
 activity of Wt1.
 CC -!- SUBUNIT: Interacts with Wt1.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- MISCELLANEOUS: 'Ciao' means 'bridge' in Chinese.
 CC -!- SIMILARITY: Contains 7 WD repeats.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; U63810; AAC24948.1; -;
 DR EMBL; AC004020; AAC23493.1; -;
 DR EMBL; BC001395; AAH01395.1; -;
 DR EMBL; BC032812; AAH32812.1; -;
 DR MIM; 604333; -;
 DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 7.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS00682; WD_REPEATS_2; 6.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Nuclear protein.
 FT REPEAT 14 53 WD 1.
 FT REPEAT 59 98 WD 2.
 FT REPEAT 103 142 WD 3.
 FT REPEAT 148 187 WD 4.
 FT REPEAT 192 231 WD 5.
 FT REPEAT 250 289 WD 6.
 FT REPEAT 301 338 WD 7.
 SEQUENCE 339 AA; 37840 MW; 63A8D8257A204FC8 CRC64;

Query Match 12.8%; Score 349; DB 1; Length 339;
 Best Local Similarity 30.4%; Pred. No. 4.2e-18;
 Matches 92; Conservative 43; Mismatches 88; Indels 80; Gaps 11;
 QY 222 VPSNKDVTS--LDWNSEGTLLATGSYDGFARIWTXKGN---LASTLGO-HKGPIFALKWN 275
 DB 11 VPAHPSRCWFLANPAGTLLASCQDRIIRINGTEGDSWICKSVLSEGHQRTVRKVAWS 70
 QY 276 KXGNFILSAGVDKTTIINDAHTGEAKQQPPFHSAPALVDVQSNNTFASCSTDMCHVCK 335
 DB 71 PCGNVLSASFPDATTCTI-----WKNQDDFECVT----- 99
 QY 336 LQDRIPIKTFQGHTEVNAIKWDPTGNLLASCDDMTLKIWMKQDN---CVHDIQAHNK 392
 DB 100 -----TLEGHENEVKSVAWAPSGNLLATCSRDKSVWVWEVEDEEYECVSLNHTQ 151
 QY 393 EIVTTIKWSPGTGNTNNPNANMLASASFDSTVRLW--DVDRGICIHITLTKHQEPVYSVAF 450
 DB 152 DVKHVVWHP-----SQELLASASYDDTVKLYREEEDDWVCCATLEGHSTVWSLAF 202
 QY 451 SPDRYILASGSPDKCVHIWNTOTGALVHSYRGTTGTFVVCVNAAGDKYGAASDGS---V 507
 DB 203 DFGGRLASCSDRTVRIWROYLPG-----NEQG--VACSGSDPSWKCI 244
 QY 508 CVL 510
 DB 245 CTL 247

Search completed: August 9, 2004, 16:46:13
 Job time : 9.66567 secs

Result No.	Score	Query		DB ID	Description
		Match	length		
1	2691	98.5	514	AAB95225	Human pro
2	2688	98.4	514	AAB07190	Human p53
3	2439	89.3	577	ADd14051	Human src
4	2342	'85.8	542	ABp41760	Human ova
5	2027	74.2	395	ABe51424	Human MDD
6	1906.5	69.8	700	ABE60376	Drosophil
7	1891.5	69.3	584	ABG21351	Novel hum
8	1142	41.8	208	AAC04385	Human pol
9	824	30.2	167	ABG21350	Novel hum
10	498.5	18.3	535	ABRS2980	Protein s
11	375.5	13.7	515	ABR53774	Protein s
12	374.5	13.7	414	ADA13321	Human int
13	374.5	13.7	514	AAR58981	WD-40 dom
14	366.5	13.4	481	ABBS9486	Drosophill
15	365	13.4	411	ABG62260	Drosophill
16	365	13.4	479	AAy79678	Drosophill
17	364.5	13.3	485	ABA68284	Amino aci
18	363.5	13.3	485	AAB68516	Human GTP
19	363.5	13.3	485	AAB92844	Human pro
20	363.5	13.3	485	ABBS7306	Novel hum
21	356.5	13.1	484	AAB68282	Amino aci
22	355	13.0	352	AAG14893	Arabidops
23	355	13.0	352	AAG48119	Arabidops
24	355	13.0	352	ABD95040	A. thalia
25	351	12.9	349	AAG48120	Arabidops

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 514 AA;

Query Match 98.5%; Score 2691; DB 4; Length 514;
 Best Local Similarity 98.6%; Pred. No. 2.5e-245;
 Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQYV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAATNQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAASQ 120
 QY 121 QGSAXNGENTANGENGAAHTIANNHTDMWEVDGVEIPSNKAVVLRGHESEVFICANPV 180
 DB 121 QGSAXNGENTANGENGAAHTIANNHTDMWEVDGVEIPSNKAVVLRGHESEVFICANPV 180
 QY 181 SLLVSGSGDSTARVNLSENSTSGPTOLVRHCIREGGQDVPSNKDVTSLDWNSECTLL 240
 DB 181 SLLVSGSGDSTARVNLSENSTSGPTOLVRHCIREGGQDVPSNKDVTSLDWNSECTLL 240
 QY 241 ATGSYDGFARVTDGNLSTLQGHKGFIFALKWNKGNFILTSGVDTKTIINDAHTGEA 300
 DB 241 ATGSYDGFARVTDGNLSTLQGHKGFIFALKWNKGNFILTSGVDTKTIINDAHTGEA 300
 QY 301 KOQPFPHAPALVDVQSNNTFASCTDMCHVCKLGODRPIKTFQGHTEVNAIKWDPT 360
 DB 301 KOQPFPHAPALVDVQSNNTFASCTDMCHVCKLGODRPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDMTLKTWSMKQNCVHDLOAHNKEIYTIKWSPTGPTNPNANMLASASF 420
 DB 361 GNLLASCSDMTLKTWSMKQNCVHDLOAHNKEIYTIKWSPTGPTNPNANMLASASF 420
 QY 421 DSTVRLWDVDRGICHTLTQKOBPVYSVAFSPDGRYLASGDFKCVHWTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTQKOBPVYSVAFSPDGRYLASGDFKCVHWTQTGALVHSY 480
 QY 481 RTGGIFPVCVNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RTGGIFPVCVNAAGDKVGASASDGSVCVLDLRK 514

RESULT 2
 ID ABO07190
 XX ABO07190 standard; protein; 514 AA.

AC ABO07190;

DT 13-AUG-2003 (first entry)

DE Human p53 modifying protein, SEQ ID 150.

KW Human; p53 modifier; cytosolic; cancer; cytosolic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.

OS Homo sapiens.
 XX WO200299122-A1.
 XX 12-DEC-2002.
 XX 03-JUN-2002; 2002WO-US017382.
 XX 05-JUN-2001; 2001US-0296076P.
 PR 10-OCT-2001; 2001US-0328605P.
 PR 15-FEB-2002; 2002US-0357253P.
 XX
 PA (EXEL-) EXELIXIS INC.
 XX
 PI Friedman L, Flowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 XX
 DR WPI; 2003-156859/15.
 DR N-PSDB; ACD13365.

Identifying modulators of the p53 pathway for use in treating apoptotic
 or cell proliferation disorders, comprises screening for agents that
 modulate activity of a human ortholog of genes that modify the p53
 pathway in *Drosophila*.

Example 2; Page 459-459; 678pp; English.

The invention relates to identifying (M1) a candidate p53 pathway
 modulating agent, by contacting an assay system comprising a purified HM
 polypeptide (human ortholog of genes that modify the p53 pathway in
Drosophila) or nucleic acid with a test agent under conditions, where but
 for the presence of the test agent, the system provides a reference
 activity, and detecting a test agent-biased activity of the assay system.
 Also included are modulating (M2) a p53 pathway of a cell (comprising
 contacting a cell defective in p53 function with a candidate modulator
 that specifically binds to a HM polypeptide comprising an HM amino acid
 sequence, where p53 function is restored; modulating (M3) a p53 pathway
 in a mammalian cell (comprising contacting the cell with an agent that
 specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
 a disease in a patient (comprising: (a) obtaining a biological sample
 from the patient; (b) contacting the sample with a probe for HM
 expression; (c) comparing the results with a control; and (d) determining
 whether the comparison indicates a likelihood disease). (M1) is useful
 for identifying modulators of the p53 pathway. A probe for HM expression
 is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
 in a patient, where the cancer has greater than 25 % expression level.
 CC Modulators identified by (M1) are useful in a variety of diagnostic and
 therapeutic applications, where disease or disorder prognosis is related
 to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
 proliferation disorders (e.g. cancer). Another two new methods (M2 and
 M3) are useful for modulating the p53 pathway of a cell, thus restoring
 the p53 function of the cell, so that the cell undergoes normal
 CC proliferation or progression through the cell cycle. (M2) and (M3) are
 also useful for treating defects in the p53 pathway such as angiogenic,
 CC apoptotic or cell proliferation disorders. The present sequence
 CC represents a human p53 pathway modifying protein

XX Sequence 514 AA;

Query Match 98.4%; Score 2688; DB 6; Length 514;
 Best Local Similarity 98.6%; Pred. No. 4.7e-245;
 Matches 507; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQYV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAATNQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAASQ 120
 QY 121 QGSAXNGENTANGENGAAHTIANNHTDMWEVDGVEIPSNKAVVLRGHESEVFICANPV 180
 DB 121 QGSAXNGENTANGENGAAHTIANNHTDMWEVDGVEIPSNKAVVLRGHESEVFICANPV 180

QY 131 SLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKDVTSLDWNSGCTLL 240
 DB 181 SLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKDVTSLDWNSGCTLL 240
 QY 241 ATGSYDGFARLWTKDGNLSTLGHKGPIFALKWKKGNFTLSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSYDGFARLWTKDGNLSTLGHKGPIFALKWKKGNFTLSAGVDKTTIINDAHTGEA 300
 QY 301 KQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDREIKTFQGHTEVNAIKWDPT 360
 DB 301 KQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDREIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSTGCTGNPNANILMASGF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSTGCTGNPNANILMASGF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQPVPYSVAFSPDGRYLASGSFDCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQPVPYSVAFSPDGRYLASGSFDCVHIWNTQTGALVHSY 480
 QY 481 RGTGGIFEVCAAGDKVKGASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCAAGDKVKGASDGSVCVLDLRK 514

RESULT 3

ADD14051
 ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

XX 01-JAN-2004 (first entry)

DT Human src biomarker polypeptide SEQ ID NO:240.

DE predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

XX WO20003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels

CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 577 AA;

Query Match 89.3%; Score 2439; DB 7; Length 577;

Best Local Similarity 86.0%; Pred. No. 2.1e-221;

Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLOESGFSGHSAFTFGIESHISOSNINGALVPPAALISIIQKLOVY 60
 DB 52 MSITSDEVNFLVRYLOESGFSGHSAFTFGIESHISOSNINGTLVPPAALISIIQKLOVY 111
 QY 61 EAEVSNEDGTLPDGRPIESLSLIIDAVMPDVVOTQCAVYDKLAQOHAHAHAHAHAHAAT-- 118
 DB 112 EAEISINEDGTVPDGRPIESLSLIIDAVMPDVVOTQCAVYDKLAQOHAHAHAHAHAHAATAA 171
 QY 119 -----NQGSAGKNGENTANGENGATIANHTDMMEVDGDIIPSNKAVVLRG 167
 DB 172 ATAAATTSAGVSHQNPCKREATVNGEENRAHSV--NNHAKPMEIDGVEIIPSKATVLRG 230
 QY 168 HESEVFICAWNPVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKD 227
 DB 231 HESEVFICAWNPVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPNSKD 290
 QY 228 VTSLDWNSGTLTATGSDYGFARINTKDNLASTLGHKGPIFALKWKKGNFTLSAGVD 287
 DB 291 VTSLDWNTNGTLTATGSDYGFARINTKDNLASTLGHKGPIFALKWKKGNFTLSAGVD 350
 QY 288 KTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDREIKTFQG 347
 DB 351 KTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDREIKTFQG 410
 QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSTGPGTN 407
 DB 411 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSTGPGTN 470
 QY 408 NPNANILMASASFDSTVRLWDVDRGICHTLTKHQPVPYSVAFSPDGRYLASGSFDCVHI 467
 DB 471 NPNANILMASASFDSTVRLWDVDRGICHTLTKHQPVPYSVAFSPDGRYLASGSFDCVHI 530
 QY 468 IWNQTGALVHSYRGTGGIFEVCAAGDKVKGASDGSVCVLDLRK 514
 DB 531 IWNQTGALVHSYRGTGGIFEVCAAGDKVKGASDGSVCVLDLRK 577

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX AC ABP41760;

DT 22-AUG-2002 (first entry)

XX Human ovarian antigen HNOX38, SEQ ID NO:2892.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory system disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 OS Homo sapiens.
 XX WO200200677-A1.
 PN 03-JAN-2002.
 XX
 XX
 PF 07-JUN-2001; 2001WO-US018569.
 XX
 XX
 PR 07-JUN-2000; 2000US-0209467P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Birse CE, Rosen CA;
 PI WPI; 2002-147878/19.
 XX N-PSDB; ABQ54837.
 DR
 DR
 XX

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX

PS Claim 11; SEQ ID NO 2892; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 542 AA;

Query Match 85.8%; Score 2342; DB 5; Length 542;
 Best Local Similarity 85.4%; Pred. No. 2.9e-212;
 Matches 434; Conservative 32; Mismatches 28; Indels 14; Gaps 2;

QY 20 GFSHSAFTGIESHSQNSINGALVPPAALISITOKGLOYVEAEVSNEDCTLEDGRPIE 79
 DB 36 GFSHSAFTGIESHSQNSINGALVPPAALISITOKGLOYVEAEVSNEDCTLEDGRPIE 95

QY 80 SLSLIDAVPDVVTQROQAYRDKLAQOHAHAARAAARAT-----NQGSACK 126
 DB 96 SLSLIDAVPDVVTQROQAYRDKLAQOHAHAARAAARATATAATTTTSAGVSHQNPKN 155
 QY 127 GENTANGEENGAAHTIANHHTDMWEVDGVEIPSNKAVLVRGHESEVFICAMNPVSDLLVS 186
 DB 156 REATVNGEENRAHSV-NNHAKPWEIDGVEIIPSSKATVLRGHESEVFICAMNPVSDLLAS 214
 QY 187 GSGDSTARINWLSNSTSGPTOLVLRHCIRREGQDVPNSKOVTSLDWNSEGTLLATGSYD 246
 DB 215 GSGDSTARINWLSNSTSGPTOLVLRHCIRREGQDVPNSKOVTSLDWNSEGTLLATGSYD 274
 QY 247 GFARIWTKDGNLSTLGHQKGFIFALKWNKKNFILSAGVDKTTIWDATGEAKQQPPF 306
 DB 275 GFARIWTKDGNLSTLGHQKGFIFALKWNKKNFILSAGVDKTTIWDATGEAKQQPPF 334
 QY 307 HSAPALDVQSNNTFASCTDMCIHVCKLGDORPIKTFQGHTEVNAIKWDPTGNLLAS 366
 DB 335 HSAPALDVQSNNTFASCTDMCIHVCKLGDORPIKTFQGHTEVNAIKWDPTGNLLAS 394
 QY 367 CSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGCTNNPNANLALASPDSTVRL 426
 DB 395 CSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSGTGCTNNPNANLALASPDSTVRL 454
 QY 427 WDVDRGICITHTLTKHQEPVYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRGCGI 486
 DB 455 WDIERGVCVHTLTKHQEPVYVAFSPDGRYLAGSFDKCVHIWNTQTGALVHSYRGCGI 514
 QY 487 FEVCWNAAGDKVGASASDGSVCVLDLRK 514
 DB 515 FEVCWNAAGDKVGASASDGSVCVLDLRK 542

RESULT 5
 ID ABP51424 standard; protein; 395 AA.
 XX
 XX ABP51424;
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 446.
 XX
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;
 KW hepatocellular carcinoma; antitumor; antiparasitic; cytostatic; anti-HIV;
 KW antiallergic; antianemic; antidiabetic; antidiarrhetic; antineoplastic;
 KW neuroprotective; antirheumatic; antiarthritic.
 OS Homo sapiens.
 XX WO2002040715-A2.
 PN 23-MAY-2002.
 XX
 XX 06-SEP-2001; 2001WO-US027628.
 XX
 XX 05-SEP-2000; 2000US-0229747P.
 PR 05-SEP-2000; 2000US-0229748P.
 PR 05-SEP-2000; 2000US-0229749P.
 PR 05-SEP-2000; 2000US-0229750P.
 PR 05-SEP-2000; 2000US-0229751P.
 PR 05-SEP-2000; 2000US-0230583P.
 PR 05-SEP-2000; 2000US-0230584P.
 PR 05-SEP-2000; 2000US-0230585P.
 PR 05-SEP-2000; 2000US-0230586P.
 PR 05-SEP-2000; 2000US-0230587P.
 PR 05-SEP-2000; 2000US-0230588P.
 PR 05-SEP-2000; 2000US-0230589P.
 PR 05-SEP-2000; 2000US-0230590P.
 PR 05-SEP-2000; 2000US-0230591P.
 PR 05-SEP-2000; 2000US-0230592P.
 PR 05-SEP-2000; 2000US-0230593P.
 PR 05-SEP-2000; 2000US-0230594P.
 PR 05-SEP-2000; 2000US-0230595P.
 PR 05-SEP-2000; 2000US-0230596P.
 PR 05-SEP-2000; 2000US-0230597P.

PR 06-SEP-2000; 2000US-0230598P.
PR 06-SEP-2000; 2000US-0230599P.
PR 06-SEP-2000; 2000US-0230610P.
PR 06-SEP-2000; 2000US-0230855P.
PR 06-SEP-2000; 2000US-0230988P.
PR 06-SEP-2000; 2000US-0230989P.
PR 07-SEP-2000; 2000US-0230951P.
PR 07-SEP-2000; 2000US-0231163P.
PR 07-SEP-2000; 2000US-0231167P.
XX (INCY-) INCYTE GENOMICS INC.
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
XX Hillman Jr, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
XX Dahl CR, Moniyama MG, Bradley DL, Rohatgi SD, Harris B;
XX Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
XX Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX WPI; 2002-527544/56.
XX N-PSDB; ABQ72641.
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
XX AIDS.
XX Claim 14; Page 578; 619pp; English.
XX The invention relates to an isolated human disease detection and
XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDDT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDDT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDDT. (I)
XX or (II) are useful for diagnosing, treating or preventing disorders
XX associated with aberrant expression of MDDT, where the disorders are
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS. Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germline gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences
XX Sequence 395 AA;
XX
XX Query Match 74.2%; Score 2027; DB 5; Length 395;
XX Best Local Similarity 98.5%; Pred. No. 1.1e-182;
XX Matches 385; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
XX
XX 1 MSISDEVNVLVYRLQESGFSHGAFTGIESHTSQSNINGALVPPAALISITKGLQYV 60
XX
XX 5 MSISDEVNVLVYRLQESGFSHGAFTGIESHTSQSNINGALVPPAALISITKGLQYV 64
XX
XX 61 EAEVSIINEDGTLFGRPIESLSLIDAVMPVQTRQAYRDKLAQOHHAAAAAATNQ 120
XX
XX 65 EAEVSIINEDGTLFGRPIESLSLIDAVMPVQTRQAYRDKLAQOHHAAAAAATNQ 124

QY 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPSKAVVLRGHESEVFI CAWNPV 180
DB 125 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPKAVVLRGHESEVFI CAWNPV 184
QY 181 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSPKNDVTSLDWNSGTL 240
DB 185 SLLASGSGDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSPKNDVTSLDWNSGTL 244
QY 241 ATGSDYGFARIWTKDGNLASTIGQHKGPFAFKWKKGNFILSAGVDKTTIWDHTGEA 300
DB 245 ATGSDYGFARIWTKDGNLASTIGQHKGPFAFKWKKGNFILSAGVDKTTIWDHTGEA 304
QY 301 KQOPFPHSAPALDVDMQSNNTFASGSTMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
DB 305 KQOPFPHSAPALDVDMQSNNTFASGSTMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 364
QY 361 GNLASCSDDMTLKIKWSMKQDNCVHDLOAHN 391
DB 365 GNLASCSDDMTLKIKWSMKQDNCVHDLOAHN 395

RESULT 6
AB860376
ID ABB60376 standard; protein; 700 AA.
XX
XX ABB60376;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 7920.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656960/75.
XX N-PSDB; ABL04479.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABLO1840-ABL16175) and the encoded proteins (ABBS57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 700 AA;
XX
XX Query Match 69.8%; Score 1906.5; DB 4; Length 700;

Best Local Similarity 54.8%; Pred. No. 7.2e-171; Matches 385; Conservative 49; Mismatches 77; Indels 191; Gaps 8;	
QY 1 MSISDEVNLFVRYLQBSGFSHSAFTGIESHSIQSNINGALVPPAALISIIQKGLQYV 60	
Db 1 MSFSSDEVNLFVRYLQBSGFLHSAYVFGIESHSIQSNINGALVPPAALTLILQKGLYT 60	
QY 61 EAEVSIINEDGTLFDGRPTIESLSLDAMPDV-----VQTRQ----- 96	
Db 61 EVEMSVGDEGV--ARPLEGJSLIDAVNPEVKPLKPIVKTPEPGVADSSAPAGGNQN 118	
QY 97 -----QAYRDKLA-----QQHAAAAAATAATNOOGSAKNGENT 130	
Db 119 NAKPEIKIEPTGVAGSAGGNKIAGSTTGTPTDQASAEVDSSGNAANAGGTACNG 178	
QY 131 ANGER----- 135	
Db 179 AGGNOASTGGNSTSTPAGGDLAAPGASQKSONSNEAGSSSGNAGNANATSTDDAASS 238	
QY 136 ---NG-----AHTIANNTDM----- 148	
Db 239 TSTNGNSTSSVEQPTSGLTPAGTGTSTSPDAAASGCASTATGSKAPSAVIRVGAQ 298	
QY 149 -----MEVDGDEVIPSNKAVILRGHESEV 172	
Db 299 GNVVQSGSSNAQSSAPSGTISSTSGGAGTPEALVPMIDENIBIPESKARVILRGHESEV 358	
QY 173 FICANVPVSLVSSGSGSTARIWNLSENSTSGPTQLVLRHCIEGGQDVPSNKNVDTSLD 232	
Db 359 FICAWNPRDLASGSGSTARIWMDSDANTNS-NQLVLRHCIOGGAEPVSNKNVDTSLD 417	
QY 233 WNSEGTILATSGYDFARITWTKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTII 292	
Db 418 WNCDSGLATSGYDGAIRWTKDGLASTLQGHKGPFPALKWKNKGNVILSAGVDKTTII 477	
QY 293 WDAHTGEAKQOPPHSAFALVDWQSNVTFASCSTDCMCIHVCKLQGRPIKTFQGHNEV 352	
Db 478 WDASTGCTQOPFAHSAFALVDWQTNQAFASCSTDQRIHVCRLGVNEPIKTFQGHNEV 537	
QY 353 NAIKWDPGTNLLASCSDDMTLKIWSMKODNCVHDLQAHNKEIYTIKWSPTGPGTNPNAN 412	
Db 538 NAIKWCFQGLLASCSDDMTLKIWSNEDRCCHDLQAHSKIEIYTIKWSPTGPGTNNTN 597	
QY 413 LMLASAFSTVRLWVDRGCIHTLTKHQBPFVYSVAFSPDGRVILASGSPDKCVHIWNTQ 472	
Db 598 LILASAFSTVRLWVDRGCIHTLTKHTEPFVYSVAFSPDGKHLASGSPDKCVHIWSTQ 657	
QY 473 TGAHVSYRGTTGGEVFCVWNAAGDKVGASADSGSVCVLDLAK 514	
Db 658 TQQLVHSYRGTTGGEVFCVWNSKGTGVKVASADSGSVFVLDLAK 699	
RESULT 7	
ABG21351	
ID ABG21351 standard; protein; 584 AA.	
XX AC ABG21351;	
XX XX	
DT 18-FEB-2002 (first entry)	
XX DE	
XX Novel human diagnostic protein #21342.	
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX KW food supplement; medical imaging; diagnostic; genetic disorder.	
XX OS Homo sapiens.	
XX XX	
XX PN WO200175067-A2.	
XX XX	
PD 11-OCT-2001.	
XX XX	
PF 30-MAR-2001; 2001WO-US008631.	
XX XX	

PR 31-MAR-2000; 2000US-00540217.	
PR 23-AUG-2000; 2000US-00649167.	
XX (HYSE-) HYSEQ INC.	
XX Dmanac RT, Liu C, Tang YT;	
XX WPI: 2001-639362/73.	
DR N-PSDB; AAS85538.	
XX New isolated polynucleotide and encoded polypeptides, useful in	
PT diagnostics, forensics, gene mapping, identification of mutations	
PT responsible for genetic disorders or other traits and to assess	
PT biodiversity.	
XX Claim 20; SEQ ID NO 51710; 103pp; English.	
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)	
XX sequences. (I) is useful as hybridisation probes, polymerase chain	
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,	
CC and in recombinant production of (II). The polynucleotides are also used	
CC in diagnostics as expressed sequence tags for identifying expressed	
CC genes. (I) is useful in gene therapy techniques to restore normal	
CC activity of (II) or to treat disease states involving (II). (II) is	
CC useful for generating antibodies against it, detecting or quantitating a	
CC polypeptide in tissue, as molecular weight markers and as a food	
CC supplement. (II) and its binding partners are useful in medical imaging	
CC of sites expressing (II). (I) and (II) are useful for treating disorders	
CC involving aberrant protein expression or biological activity. The	
CC polypeptide and polynucleotide sequences have applications in	
CC diagnostics, forensics, gene mapping, identification of mutations	
CC responsible for genetic disorders or other traits to assess biodiversity	
CC and to produce other types of data and products dependent on DNA and	
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic	
CC amino acid sequences of the invention. Note: The sequence data for this	
CC patent did not appear in the printed specification, but was obtained in	
CC electronic format directly from WIPO at	
CC ftp.wipo.int/pub/published_pct_sequences	
XX	
QY Sequence 584 AA;	
Best Match 69.3%; Score 1891.5; DB 4; Length 584;	
Best Local Similarity 64.6%; Pred. No. 1.4e-169;	
Matches 368; Conservative 34; Mismatches 39; Indels 129; Gaps 4;	
QY 16 LQESGFSHSAFTGIESHSIQSNINGALVPPAALISIIQKGLQYVEAEVSIINEDGTLFDG 75	
Db 62 LEVKGFSHSAFTGIESHSIQSNINGTLVPPSALISILQKGLQYVEAEVSIINEDGTLFDG 121	
QY 76 RPTESLSLDAMPDVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 124	
Db 122 RPIESLSLVAVIPDVQVQNRQAFGEKLTQOASAAATASMAKATMTPAISQOQNP 181	
QY 125 KNGENTANGEANGAHTIANNHTDMMEDVDGVEIPSNKAVILRGHESEVFIICAWNPSDLL 184	
Db 182 KQREATVNGEANGAHEI-NNHSPKWEIDGVEIPENKATVLRGHESEVFIICAWNPSDLL 240	
QY 185 VSGSGDSTARIWNLSENSTSGPTQLVLRHCIEGGQDVPSNKNVDTSLDWSNSETLLATGS 244	
Db 241 ASGSGDSTARIWNLSENSTSGPTQLVLRHCIEGGQDVPSNKNVDTSLDWSNSETLLANGS 300	
QY 245 YDGFARINTKDNGLASTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEAKQOF 304	
Db 301 YDGFARINTWTE----- 310	
QY 305 PHSAFALVDWQSNVTFASCSTDCMCIHVCKLQGRPIKTFQGHNEV----- 349	
Db 311 ---NAPALVDWQNNMTFASCSTDCMCIHVCRUGCDHPKATPFQGHNEVCTCIBSICFAGGLR 367	
QY 350 -----NEVNAIKWDPT 360	
Db 368 KLTMTTEGRLRPKTCFSDGALLPPAGRRPHLLTGPDIFFKRLSALFQNEVNAIKWDPS 427	

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 167 AA;

Query Match 30.2%; Score 824; DB 4; Length 167;

Best Local Similarity 86.8%; Pred. No. 2,1e-69;

Matches 145; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

250 RIWTKDGNLSTLGHKGFIPALKNKGNFILSAGVDKTTIWDATGEAKQFPFHA 309

1 RIWTEGNLAITLCHKGFIPALKNKGNFVLSAGVDKTTIWDATGEAKQFPFHA 60

310 PALDVDMQNNFTASCSTDMCIHVCKLGODRPIKTFQHTNEVNAIKWDPTGNLLASCSD 369

61 PALDVDMQNNFTASCSTDMCIHVCKLGODRPIKTFQHTNEVNAIKWDPTGNLLASCSD 120

370 DMTLKWSMKQNCVHDLQAHNKETIYTIKWSPTGPTGNPNANMLA 416

121 DMTLKWSMKQNCVHDLQAHNKETIYTIKWSPTGPTGNPNANMLA 167

RESULT 10

ABRS2980

ID ABR52980 standard; protein; 535 AA.

AC ABR52980;

DT 20-JUN-2003 (first entry)

XX Protein sequence #SEQ ID 825.

DE Multi-protein complex; eukaryote; drug target; diagnosis.

KW Saccharomyces cerevisiae.

OS

XX EP1258494-A1.

XX 20-NOV-2002.

XX 20-DEC-2001; 2001EP-00130253.

XX 15-MAY-2001; 2001EP-00111774.

XX (CELL-) CELLZONE AG.

XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzioch M, Schultz JD, Superti-Furga GD;

DR WPI; 2003-250078/25.

DR N-PSDB; ACC61022.

XX New isolated protein complexes useful for diagnosing a disease or

PT disorder, or as a target for an active agent of a pharmaceutical,

PT preferably a drug target in the treatment or prevention of disease or

PT disorder.

XX Disclosure; SEQ ID NO 825; 17pp + Sequence Listing; English.

XX The invention relates to multiprotein complexes from eukaryotes. Proteins

CC of the invention and DNA sequences encoding them are given in records

CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are

CC obtainable by using a protein as a bait and isolating the set of proteins

CC which is attached thereto from cells. Such protein complexes may comprise

CC up to 30 distinct proteins. Protein complexes of the invention are useful

CC for diagnosing a disease or disorder, or as a target for an active agent

CC of a pharmaceutical, preferably a drug target in the treatment or

CC prevention of a disease or disorder. Note: The sequence data for this

CC patent is not represented in the printed specification, but is based on

CC sequence information supplied by the European Patent Office. The complete

CC document is available on CD-ROM

XX Sequence 535 AA;

Query Match 18.3%; Score 498.5; DB 6; Length 535;

Best Local Similarity 25.4%; Pred. No. 8.7e-38;

Matches 143; Conservative 101; Mismatches 239; Indels 79; Gaps 16;

1 MSISDEVNLFVRYLQESGSHSAFTFGIESHSQSNN-CALVPPAALISIIQKQY 59

1 MSITSEELNYLWRYCOEKGHEVSALALQDETRVLEFDEKYEKHEIPLGLTLNVLVQGIY 60

60 VRAEVSINEDG---TLFGRPIESLSLDVNPDPVVTQRAYRKLAQQHAAAAA 116

61 TESELMVDKSGDISALNEHLSDFNLVQALQID-----KEKFFE--ISSEGREFTL 109

117 ATNCOGSAGKNGENTANGEANGAHTIANNHDTMW-EVDGDVEIPSKAVVLRGHESEVFC 175

110 ETNSE-SNKAAGDASTVERETOEDDTNSIDSSDDLDGFKI--LKEIV--KLDNIYSS 163

176 ANNPVSD-LLVSGSGDSTARIWNLSNSTSGPTQ-----LVLRHCIREGGODVPSNKDV 228

164 TWPPLDESILAYCEKNSVARLARIVETDQEGKKYWKLTIIAELRHPFALSASSGKTTNQV 223

229 TSLDWNSECTLLATGSDGFARIWTKDGNLASTLGHKGFIPALKNKGNFILLSAGVDK 288

224 TCLAWSHDGNISIVTGVENGELRLNKTGALLNVLNFRHAPIVSVRWKNDGTHIISNDVEN 283

289 TTIWDAHTGEAKQOFPF-----HSAP---ALDVDMQNNFTASCSTDMCIHV 333

284 VTILMNVISGTVMQHPFELKTCGSSINAENHSGDGLGVDEWVDDDKFVIPGKGAIFV 343

334 CKLGQDRPIKTFQHTNEVNAIKWDPTGNLLASCDDMTLKWSMKQNCVHDLQAHNKE 393

344 YQITEKTPTKLIGHHGPISVLEFNDTNKLLSASDDGTLRWGGNGNSQNCFFYGHQS 403

394 IYTIKWSPTGPTGNPNANMLASASFDSTVLRLWVDVDRGICITHTLTKHQEPYVSVAFSD 453

404 IVSASWV-----GDDKVISCSMDGSLVLSLQNTLLALSIVDGVPIFAGRIQSD 453

454 GRYLASGSPDKCVHIW-----NTQTGAL-----VHSYRGTTGIFVC 490

454 GQKYAVAFMDGVQNVYDLKLNKSKRSRYGNRDGILNPLPIPLYASYOSSQNDYIFDLS 513

491 WNAAGDKVGA--SASDGSVCVL 510

514 WNCAGNKISVAYSLOEGSVVAI 535

RESULT 11

ABRS3774

ID ABR53774 standard; protein; 515 AA.

XX ABR53774;

XX 20-JUN-2003 (first entry)

XX Protein sequence #SEQ ID 2413.

XX Multi-protein complex; eukaryote; drug target; diagnosis.

XX Saccharomyces cerevisiae.

XX EP1258494-A1.

XX 20-NOV-2002.

XX 20-DEC-2001; 2001EP-00130253.

XX 15-MAY-2001; 2001EP-00111774.

XX (CELL-) CELLZONE AG.

XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzioch M, Schultz JD, Superti-Furga GD;

DR WPI; 2003-250078/25.

Query Match 13.7%; Score 374.5; DB 6; Length 414;
 Best Local Similarity 26.7%; Pred. No. 3.2e-26;
 Matches 111; Conservative 55; Mismatches 165; Indels 85; Gaps 13;

QY 83 LIDAVMPVQTRQAYRDKLAQQ-----HAAAAAATAAQQGSA-KNGE-- 128
 DB 54 LLTASRTQVKLLIQRLQKLGQNSHTFYLKVKLKAHILPLTNVALNKSGSCFITGSDYD 113
 QY 129 -----NTANGENGCAHTIANNHTDMEVDGVDVIPSNAVVLSCHESEVEFICAW-NPVS 181
 DB 114 RTCKLWDITASGEELN-----TLGHRNVVYATAFNPYG 147

QY 182 DLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLA 241
 DB 148 DKIATGSDFKTKLWSVETG-----CYHTR-----GHTAEIVCLSFNPQSTLVA 193

QY 242 TGSYDGFARIWTKGNLASTIGQHKGPITFALKWKNKGNFILSAGVDKTTIWDHATGEAK 301
 DB 194 TGSMDTTAKLWDIQNGEELTLRGHSABEISLSFNTSGDRIITGTFDHTVVVWMDADTGRKV 253

QY 302 QQFPFSA--PALVDVQSNNTFASCS---TDMCHVCKL---GQDRPIKTFQGTNEVN 353
 DB 254 NILIGHCAEISSANW-----DCSLILTGSMCKTKLWDATNGKCVALTGHDDDEIL 306

QY 354 AIKWDPTGNLLASCDMDTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPGTGPTNPNANL 413
 DB 307 DSCFDYTKGLIATASADGTAFISAAATKRAKIAKLEHGEHEISKISF-----NPQGNH 358

QY 414 MLASFSFTVRLWDVDRGICIHILTKEQEPVYSVAFSPDGRYLASGDFKCVHIW 469
 DB 359 LLTGSS-DKTAIRWDAQTGQCLQVLEHTDEIFSCAFNYKGNIVITGSKONTCRIW 413

RESULT 13

ID AAR85881 standard; protein; 514 AA.
 XX AAR85881;
 AC AAR85881;
 XX AAR85881;
 DT 13-SEP-1996 (first entry)
 DE WD-40 domain-contg. YCW2 protein.
 XX WD40 repeat region; beta-transducin; protein-protein interaction; drug;
 KW intracellular signalling; protein kinase C; homology; motif; modulator;
 KW receptors of activated protein kinase; enzyme activity; isozyme; human.
 XX Synthetic.
 OS
 XX WO9521252-A2.
 FN
 XX 10-AUG-1995.
 PD
 XX 31-JAN-1995; 95WO-US001210.
 PF
 XX 01-FEB-1994; 94US-00190802.
 PR
 XX (STRD) UNIV LELAND STANFORD JUNIOR.
 PA
 XX Mochly-Rosen D, Ron D;
 PI
 XX WPI; 1995-283772/37.
 DR
 XX New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the
 PT activity of a protein, eg. protein kinase C, which interacts with a
 PT protein contg. a WD-40 region.
 XX
 PS Example 5; Page 168-170; 351pp; English.
 XX
 CC Proteins AAR85851-92 are protein which contain at least one WD-40 (also
 CC called beta-transducing homologous) amino acid repeat motifs. The WD-40
 CC regions are involved in protein-protein interactions between proteins

CC involved in intracellular signalling. An example of such an interaction
 CC is between protein kinase C and receptors of activated protein kinase
 CC (RACK), esp. RACK-1 (AAR85850). Proteins AAR85851-82 were isolated based
 CC on homology with beta-transducin, whereas proteins AAR85882-92 were
 CC isolated based on homology with the WD-40 consensus sequence (AAR85893).
 CC The proteins were used to construct the peptides AAR84928-R85063 and
 CC AAR85786-R85842. The peptides can be used to identify target proteins
 CC contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of
 CC proteins involved in protein-protein interaction and to screen for drugs
 CC that will affect protein-protein interaction involving WD-40 domains
 XX
 XX Sequence 514 AA;

Query Match 13.7%; Score 374.5; DB 2; Length 514;
 Best Local Similarity 23.6%; Pred. No. 4.5e-26;
 Matches 125; Conservative 85; Mismatches 167; Indels 153; Gaps 21;

QY 38 NINGALVPPAALISIIQKLVQVEAEVINEGDTI-----FDGR-----PIESLSLIDA 86
 DB 44 NVGGALRVPCA---ISEKLEELNLQNGTSDDPVYTFSTCTIOGKKASDPVKTIIDITDN 100
 QY 87 VMPDVVQTRQAYRDKLAQOHA 146
 DB 101 LVSSLIKPGYNSTEQIITLLYTPRAVFKVKPYTRSSA----- 138

QY 147 DMMEVDGVEIIPSNKAVLVRGHESEVEFICAMNP-VSDLLVSGSGDSTARIWNLSNSTSG 205
 DB 139 -----IAGHGSTILCSAFAPHTSSRMVTGAGDNTARIWDC---DTQT 177

QY 206 PTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATGSDGFARIW-TKQGN-LASTLG 263
 DB 178 PMHTLKGH-----YNVLCVSNPSDGEVATGSMNTILNDPKSGQCLGDALR 226

QY 264 QHKGPITFALKWN-----KQGN--FILSAGVDKTTIWDHATGEAKQFPFHSAPALVDW 316
 DB 227 GHSKWITSLSWEPHILVKPGSKPRLLASSSKOGTIKINDTVS----- 267

QY 317 QSNNTFASCSGDMCHIVCKLGQDRPIKTFQGTNEVNALKWDPTGNLLASCDMDTLKIWI 376
 DB 268 -----RVCQY-----TMSGHTNSVCVKWGGQG-LLYSGSHDRTVRW 304

QY 377 SM-KQDNCVHDLQAHNKEI-----YTIK---WSPTG---PGT----- 406
 DB 305 DINSQGRGINILKSHAHVWNLSTLDYALRICAFDHTGKXKPEEAKKALENYEKIC 364

QY 407 -NNPNANILMASFSDSTVRLWDVDRGI-CIHTLTKHQEPVYSVAFSPDGRYLASGSDK 464
 DB 365 KKNGNSEEMVMTASDDYTMFLNPLKSTKPIARMTGHQKLVNHNHAFSPDGRYIVSASFDN 424

QY 465 CVHIWNTQTGALVHVSVRG-TGGIFEVCVNAAGDKVGSASDGSVCVLDLR 513
 DB 425 SIXLWDGRDGKEISTFRGHIAVSYQVAMSSDCELLVSCSKDTTLKVDVR 474

RESULT 14
 ABB59486
 ID ABB59486 standard; protein; 481 AA.
 XX ABB59486;
 AC ABB59486;
 XX ABB59486;
 DT 26-MAR-2002 (first entry)
 DE
 XX Drosophila melanogaster polypeptide SEQ ID NO 5250.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW Drosophila; pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX

```
PF 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL03589.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 5250; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 481 AA;
XX
XX Query Match 13.4%; Score 366.5; DB 4; Length 481;
XX Best Local Similarity 27.5%; Pred. No. 2.3e-25;
XX Matches 128; Conservative 61; Mismatches 175; Indels 101; Gaps 21;
XX
QY 115 AATNQ-----GSAKNGENTAN-----GREGAHTIANNHTDMEVDCG--VEIPSNKA 162
Db 35 AGITTQGLIGNALLKNEEATPYLFVGEDEIKKSLEDT-LDLASVDNTENVIDIYVQPQ 93
QY 163 VLNR-----CHESEVFICAWNPVSDLLVSGSGDSTARINWLSN-----STSGP 206
Db 94 AVFKVRPVTRCTSSMPGHAENVSLNPSPDGAHLASGSGTTVRLDNLNTETPHFTCTGH 153
QY 207 TQLVLRHCIR-----EGQD-----VPSNKDVTSLDW-----N 234
Db 154 KQWVL--CVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEPYHRD 211
QY 235 SEGTLATGSDGPARTW-TKDGNIASLTGQHGPIFALKWKNKGNFILLSAGVDKTTIIV 293
Db 212 PECKLASAGDGCRIWDVLGQCLMNIAGHTNAVTAVERWGAG-LIYTSSKDRIVKMW 270
QY 294 DAHTGEAKQFPFPHSAPALDVMQSNNTFASCTDMCIHCKLGQDRPIK-TFQGH----- 348
Db 271 RAADGILCRTFSGHA-----HWNN--TALSTD---YVLRTPGFPHVDRSKSHLSLS 318
QY 349 TNEV-----NAIKWDPGNLLASCSDDMTLKIMSKQDNCVHDLOAHNKEIYTIKWS 400
Db 319 TEELQESALKRYQAVCPDEVESLV-SCSDNTLYLWERNQKCVERTMGTQNVNDVKYS 377
QY 401 PTGPGTNNPNNANMLASAFDSTVRLMDVDRGICHTLTIKHQPVPVSVAFSPDGRVLASG 460
Db 378 PDVK-----LIASAFDKSVRLWRASDGGYATFRGHVQAVYTVANSADSRLLVSG 428
QY 461 SFDKCVHNTQTGALVHSYRG-TGGIFEVCAWNAAGDKVQASAD 504
Db 429 SKDSTLKVMSYQTKLAQELFGHADEVFGVDWAPDGSRVASGGKD 473
XX
XX RESULT 15
XX ID ABB62260
XX AC ABB62260;
XX ABB62260;
XX ABB62260;
```

```
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 13572.
XX
XX Drosophila; developmental biology; cell signaling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06363.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 13572; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 411 AA;
XX
XX Query Match 13.4%; Score 365; DB 4; Length 411;
XX Best Local Similarity 22.7%; Pred. No. 2.5e-25;
XX Matches 110; Conservative 72; Mismatches 155; Indels 148; Gaps 12;
XX
QY 6 DEWNLVYRYLOESGFSHSAFTGIESHS---QSNINGALVPPAALISIIQKGLQYVEA 62
Db 11 EELNQAIADYIGNSGYADSLSTFRKADLSTVEKFKGLLEKKWTSVIRLQKKWLEEA 70
QY 63 EYSINEDGTLFDGRPIESLSLIDAYMPDVVQTRQOAYRDKLAQHAHAATAATNQO 122
Db 71 KL-----TEAEK 77
QY 123 SAKNGENTANGENGAAHTIANNHTDMEVDCGVEIPSKAVVLRGHESEVFICAWNPVSD 182
Db 78 EVIEGAPTCKKPTGSEW-----IPRPPK-FSLTGHRSATIRVIFHPIFA 121
QY 183 LLVSGSGDSTARINWLS-----ENSTSGTQLVLRHCIREGGQDVPSPKNDVTSLDWNS 238
Db 122 LMVSASEDATIRIDFETGETEYERSLKGHDTSV-----QDVA-----FDAQK 163
QY 239 LIATGSDYDGFARIW---TKDGNLASTLGQHGPIFALKWKNKGNFILLSAGVDKTTI 296
Db 164 LLASCSADLSIKLWDFQQSCEYCIKTMGHGHDHNVSSVAFVDPAGDYVLSRDTIKKWE 223
QY 297 TGEAKQFPFPHSAPALDVMQSNNTFASCTDMCIHCKLGQDRPIKTFQGHNEVNAIK 356
Db 224 TG-----YC-----VKTYTGHEWVRVVR 242
```

```

QY 357 WDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSP-----TGPCTNN 408
Db 243 VHISSIFATCSNDQIRVWLNTSKDCKVELRDREHTVECIAMAPAAAAAAGADN 302
QY 409 PNAN---LMLASASFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKC 465
Db 303 KKGHHQGPFLASGSRDKTIRIWDVSVGLCLLTLGHDNWWVRGLAFHPGGKYLVSASDDKT 362
QY 466 VHIWN 470
Db 363 IRVWD 367

```

Search completed: August 9, 2004, 16:45:35
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:43 ; Search time 13.3333 Seconds
(without alignments)
3708.183 Million cell updates/sec

Title: US-09-987-701-2
Perfect score: 2726
Sequence: 1 MSISDEVNLYRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	502.5	18.4	535	2 S48268	probable membrane
2	479	17.6	1526	2 AC2239	WD-40 repeat prote
3	463.5	17.0	1258	2 AI2155	WD-repeat protein
4	460	16.9	1356	2 T18521	beta transducin-li
5	445.5	16.3	1683	2 AF2071	WD-40 repeat prote
6	439.5	16.1	1189	2 AI2433	WD-repeat protein
7	439	16.1	1227	2 AE1810	WD-40 repeat prote
8	437	16.0	1708	2 AE1866	WD-40 repeat prote
9	414	15.2	1711	2 AD1842	WD-40 repeat prote
10	412.5	15.1	934	2 AG1889	WD-40 repeat prote
11	411.5	15.1	1747	2 AC1842	WD-40 repeat prote
12	390.5	14.3	564	2 T40893	WD repeat protein
13	388.5	14.3	1693	2 S76086	beta transducin-li
14	382.5	14.0	676	2 AB2195	hypothetical prote
15	382	14.0	1551	2 AB2410	WD-repeat protein
16	371.5	13.6	502	2 T41148	trp-asp repeat con
17	368.5	13.5	304	2 AG1837	WD-40 repeat prote
18	368.5	13.5	1189	2 AH2154	WD-repeat protein
19	364.5	13.4	515	2 S19487	hypothetical prote
20	360	13.2	265	2 AF1890	WD-repeat protein
21	350.5	12.9	677	2 AE1861	serine/threonine k
22	348.5	12.8	437	2 S05357	hypothetical prote
23	344.5	12.6	410	2 S48052	platelet-activatin
24	344.5	12.6	559	2 AB2202	hypothetical prote
25	339.5	12.5	786	2 AG2375	WD-40 repeat-prote
26	337	12.4	409	2 S36113	LIS-1 protein - hu
27	335	12.3	333	2 C85034	probable WD-repeat
28	333	12.2	323	2 T02617	hypothetical prote
29	330.5	12.1	473	2 T33805	hypothetical prote

30	329	12.1	798	2 S34023	TATA box-binding p
31	326.5	12.0	787	2 T00798	hypothetical prote
32	324	11.9	342	2 AE2490	WD-repeat protein
33	323.5	11.9	777	2 T41075	hypothetical WD-te
34	321	11.8	376	2 T19266	hypothetical prote
35	320.5	11.8	589	2 AG2400	WD-repeat protein
36	319.5	11.7	606	2 T08180	PF20 protein, micr
37	317.5	11.6	579	2 T22703	hypothetical prote
38	310.5	11.4	704	2 S33263	transcription init
39	309.5	11.4	357	2 AI2099	WD-40 repeat prote
40	309	11.3	1049	2 T42045	beta transducin-li
41	309	11.3	1191	2 S76414	beta transducin-li
42	308	11.3	640	2 S49932	MET10 protein - ve
43	305.5	11.2	317	2 T48032	WD-40 repeat regul
44	304	11.2	605	2 T38932	probable sulfur me
45	300	11.0	701	2 T16607	hypothetical prote

ALIGNMENTS

RESULT 1

S48268
probable membrane protein YBR103w - Yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR0832
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S48268; S45971; S44683
R;Manhaupt, G.; Stucka, R.; Ehnlie, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357; PMID:7900426
A;Accession: S48268
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-535 <MAN>
A;Cross-references: EMBL:X78993; NID:G476045; PIDN:CAA55606.1; PID:G476059
R;Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45971
A;Molecule type: DNA
A;Residues: 1-535 <FE2>
A;Cross-references: EMBL:Z35972; MIPS:YBR103W
C;Genetics:
A;Gene: SGD:SIF2
A;Cross-references: SGD:S0000307
A;Map position: 2R
C;Superfamily: unassigned WD repeat proteins; WD repeat homology
C;Keywords: transmembrane protein
F;355-388/Domain: WD repeat homology <WD1>
F;397-429/Domain: WD repeat homology <WD2>
F;434-450/Domain: transmembrane #status predicted <TMM>

Query Match 18.4%; Score 502.5; DB 2; Length 535;
Best Local Similarity 24.9%; Pred. No. 3.3e-28;
Matches 142; Conservative 100; Mismatches 232; Indels 97; Gaps 15;

QY	1	MSISDEVNLYRYLQESGFHSA-----FTFGIKSHISQSNINGALVPPAAL	49
Db	1	MSITSEELNYLIWRYCQEMGHEVSALALQDETRVLEFDEKYKEHI-----PLGTL	50
QY	50	ISLIQKGQYVAEVSINEDG---TLFGCRPIESLSDAVMPDVVQTRQQAYRKLQAQ	106
Db	51	VNLVQRGILYTESLMDVDSKGDISALNEHLSDEFNLVQALQID-----KEKFP	101
QY	107	QAAAAAASQCSAKNGENTANGENGAHTIANHHTDMMEVDGVEIPPNKAVLVR	166
Db	102	SSEGRFTLETNSKNKAGEDGASTVERTEQDDTNSISSD--DLDGFVKI--LKEIV--	155
QY	167	GHESVFTCAWNPVSD--LLAGSGDSTARINWLSNENSTSGSTQ-----LVLHRCIR	219
Db	156	-KLDINVSSTWNPFLDESILAYGKNSVARLARIVETDQEGKYWKLTITIAELRHPF	214

```

220 QDVPSNKDVTSLDWNSEGLLATGSDYGFARIMTKDGNLSTLQGHKGPFFALKWKNKGN 279
215 SSGKITNOVTCLAWSHDGNISVTVGENGELRLWNKTGALLNVLNFRAPIVSVKWKDGT 274
280 FILSAGVDKTTIWDHAHTGEAKQFPP-----HSAP-----ALDWDWQSNNTFAS 324
275 HISMDEVNTILMNVISGTVMQHFELKETGSSINAENHSGDGLGVDVWVDDXKFXI 334
325 CSTMCIHVKLGQDRPKITFGQHTNEVNAIKWDPNTGNLLASCDDMTLKIWSMKQNCV 384
335 PGPKAIFVYITTEKTPGKLIHGHPISVLEFNDTKLLSASDDGTLRIWHGNGNSQ 394
385 HDLQOHKEIYTIKWSPTGPTNNPNMLMLASAFDSTVRLWDVDRGICHTLTKEQEP 444
395 NSFYHSGSIVASVW-----GDDKVISCMDGVSRLWSLQNTLLALSIVDGV 444
445 VYSVAFSPDGRYLASGSPDKCHW-----NTQTAL-----VHSYR 481
445 IFAGRISQGGKYAVAFMDGQVNVYDLKLNKSKRSLSYGNRDLNPLPIPLVASYQSSQ 504
482 GTGGIFFEVCMNAAGDKVGA--SASDGSVCVL 510
505 DNDYIFDLSWNCAGNKISVAISLQEGSVVAI 535

RESULT 2
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075165.1; PID:g17132599; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 17.6%; Score 479; DB 2; Length 1526;
Best Local Similarity 29.8%; Pred. No. 6.8e-26;
Matches 117; Conservative 76; Mismatches 149; Indels 50; Gaps 10;

QY 136 NGAHTIANHDMEDVDGVEIPNKAV-VLGEHSEVFCIAWNPVSDLLASGSDSTAR 194
1127 NGV-TLANGSQIVRLWD--ISSKKCLYTLQHTNWNVNAVAFSPDGTASGSDQTVR 1183
195 IWNLSGNS-----TSGSTQLVLEHCHIREGGQDVPSNKDVTSLDWNSEGLLATGSDYGFAR 250
1184 LWDISSKCLYTLQHTSW-----VNSVFNPDGSTLASGSDQTVR 1225
251 IW-TKDNGLASTLQGHKGPFFALKWKNKGNFILSAGVDKTTIWDHAHTGEAKQFPHSA 309
1226 LWEINSSKCLCTFOGHTSWNSWFWNPDPGSLASGSDKTVRLWDISSKCLHTFQGH- 1284
310 PALDWDWQSNNTF-----ASCSTDMCHVCKLQDRPKITFGQHTNEVNAIKWDPNTGN 362
1285 -----NWNVSVAFNPDGSLASGSDQTVRLWEISSKCLHTFQHTSWVSSVTFSPDGT 1339
363 LLASGDDMTLKIWSKQNCVHDLQOHKEIYTIKWSPTGPTNNPNMLMLASAFDS 422
1340 MLASGSDQTVRLWISGSECLYTLFHTNWNVSVAFSPDGT-----AILASGSDQ 1390
423 TVRLWDVDRGICHTLTKEQEPVYSVAFSPDGRYLASGSDKCHWIMNTQTGALVHSYR 482

Db 220 QDVPSNKDVTSLDWNSEGLLATGSDYGFARIMTKDGNLSTLQGHKGPFFALKWKNKGN 279
215 SSGKITNOVTCLAWSHDGNISVTVGENGELRLWNKTGALLNVLNFRAPIVSVKWKDGT 274
280 FILSAGVDKTTIWDHAHTGEAKQFPP-----HSAP-----ALDWDWQSNNTFAS 324
275 HISMDEVNTILMNVISGTVMQHFELKETGSSINAENHSGDGLGVDVWVDDXKFXI 334
325 CSTMCIHVKLGQDRPKITFGQHTNEVNAIKWDPNTGNLLASCDDMTLKIWSMKQNCV 384
335 PGPKAIFVYITTEKTPGKLIHGHPISVLEFNDTKLLSASDDGTLRIWHGNGNSQ 394
385 HDLQOHKEIYTIKWSPTGPTNNPNMLMLASAFDSTVRLWDVDRGICHTLTKEQEP 444
395 NSFYHSGSIVASVW-----GDDKVISCMDGVSRLWSLQNTLLALSIVDGV 444
445 VYSVAFSPDGRYLASGSPDKCHW-----NTQTAL-----VHSYR 481
445 IFAGRISQGGKYAVAFMDGQVNVYDLKLNKSKRSLSYGNRDLNPLPIPLVASYQSSQ 504
482 GTGGIFFEVCMNAAGDKVGA--SASDGSVCVL 510
505 DNDYIFDLSWNCAGNKISVAISLQEGSVVAI 535

RESULT 3
AI2155
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AI2155
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2155
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1258 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074499.1; PID:g17131893; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2800

Query Match 17.0%; Score 463.5; DB 2; Length 1258;
Best Local Similarity 31.8%; Pred. No. 6.7e-25;
Matches 103; Conservative 44; Mismatches 100; Indels 77; Gaps 7;

QY 165 LRGEHSEVFCIAWNPVSDLLASGSDSTARINWLSNENSTSGSTQLVLRHCIREGGQDVPS 224
974 LEHTDFTIIGIAFSPDSQTLASASTDSSVRLWNI---STGQCFQILLH----- 1019
225 NKDVTSLDW-----NSEGLLATGSDYGFARIMTKDGNLSTLQGHKGPFFALKWKNK 277
1020 -----TDWYVAVFHPQGGKIATGSADCTVKLWNIISGQCLKLTSEHSDKTLGNWSPD 1073
278 GNFILSAGVDKTTIWDHAHTGEAKQFPHSAFALDWDWQSNNTFASCTDMCHVCKLG 337
1074 GOLLASASADQSVRLWD-----CCTGRCVGI----- 1099
338 QDRPKITFGQHTNEVNAIKWDPNTGNLLASCDDMTLKIWSMKQNCVHDLQOHKEIYTI 397
1100 -----LRGHSNRVYSAIFSPNGEIIATCSTDQTVKIWDWQGGKCLKTLTGHNTWVFDI 1152
398 KWSPTGPTNNPNMLMLASAFDSTVRLWDVDRGICHTLTKEQEPVYSVAFSPDGRYL 457
1153 AFSPDGT-----ILASAGHDQTVRIWDVNTGKCHHCICHTLHVSVAFPDGEV 1203
458 ASGSFDCVCHWIMNTQTGALVHSYR 481
1204 ASGSQDQTVRIWNVKTEGCLQILR 1227

RESULT 4
T18521
beta transducin-like protein - Podospora anserina
C:Species: Podospora anserina
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18521
R:Saue, S.; Turcq, B.; Begueret, J.
Gene 162, 135-139, 1995
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anseri
A:Reference number: Z18944; MUID:96009891; PMID:7557402
A:Accession: T18521
A>Status: preliminary; translated from GB/EMBL/DDSDJ
A:Molecule type: DNA
A:Residues: 1-1356 <SAU>
A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1
C:Genetics:
```

A;Gene: het-e1
A;Introns: 761/3

Query Match
Best Local Similarity 16.9%; Score 460; DB 2; Length 1356;
Matches 124; Conservative 73; Mismatches 178; Indels 70; Gaps 12;

QY 78 IESLSLDVAMPDVVQTRQYQKLAQQQAAAAAASQGGAKNGE-----NT 130
DB 820 ISISVVEAENACTQT-----LEHGSSVLSVAFSADQQRVASGSDDKTIKIWD 870
QY 131 ANGENGAHTIANNHTDMEVDGVEIPPNKAVLVRGHESEVFIKAWNFVSDLLASGSD 190
DB 871 ASG--TGTQT-----LEHGSSVLSVAFSADQQRVASGSD 904
QY 191 STARIWLSNLSGSG--TOLVLRHCIREGGQDVPSNKDVLSLWNSGTTLLATGSDGPA 249
DB 905 KTIKW-----DASGTCQTGLEH-----GGR-----VOSVAFSPDQQRVASGSDHTI 949
QY 250 RIW-TKDNGLASTLGHQKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEAKQPPFHS 308
DB 950 KINDAASGTCQTGLEHGSSVLSVAFSPDQQRVASGSDKTIKIWDTASGTCQTGLEHG 1009
QY 309 APALVDWQ--SNTFASCSTDMCIHVCKLQGDRIKTFQHTNEVNAIKWDPTCNLLASC 367
DB 1010 GSVMSVAFSPDQQRVASGSDKTIKIWDTASGTCQTGLEHGGSVWSVAFSPDQQRVASG 1069
QY 368 SDDMTLKIWSMKQNCVHDLQOHNKIYTIKWSPTGPTNNPNANMLASASFDSTVRLW 427
DB 1070 SDDHTIKIWDVSGTCQTGLEHGSDSVWSVAFSPDQ-----RVASGSDIGTIKIWI 1120
QY 428 DVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHVSRTGG-I 486
DB 1121 DAASGTCQTGLEHGGSVWSVAFSPDQQRVASGSDKTIKIWDASGTCQTGLEHGGSV 1180
QY 487 FEVCNNAAGDKVGASASDGSVCVLD 511
DB 1181 QSVAFSPDQQRVASGSDKTIKIWD 1205

RESULT 5
AF2071

WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AF2071
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1683 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA073823.1; PID:gl7131215; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1214

Query Match
Best Local Similarity 16.3%; Score 445.5; DB 2; Length 1683;
Matches 109; Conservative 69; Mismatches 165; Indels 40; Gaps 6;

QY 124 AKNGENTANGENGGAHTIANNHTDMEVDGVEIPPNKAVLVRGHESEVFIKAWNFVSD 183
DB 1290 SSDKALASASRDNTIKLWNRHGIELE-----TFTGSHGGVAVNLFPSDNI 1336
QY 184 LASGSGSTARIWLSNLSGSGTQLVLRHCIREGGQDVPSNKDVLSLWNSGTTLLATG 243
DB 1337 IASASLNTIKLWNRPLISPL-----EVLNAGSGVAVSFLHDSGIATA 1381

QY 244 SYDGFARIW-TKDNGLASTLGHQKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEAKQ 302
DB 1382 GADGNIQLWHSQDSGLLXLPNKA-IYGISFTPOGLIASANADKTVKIWRVROGKALK 1440
QY 303 QPFFHSAPALVDWQ--SNTFASCSTDMCIHVCKLQGDRIKTFQHTNEVNAIKWDPTG 361
DB 1441 TLIGHDNEVKNVNFSPDGTTLASASRDNTIKLWNSVDSGKFKTKLKGHTDEVFWVSFSDG 1500
QY 362 NLASCSDDMTLKIWSMKQNCVHDLQOHNKIYTIKWSPTGPTNNPNANMLASASFD 421
DB 1501 KIIASASADKTIIRLWDSFGNLIKSLPAHNDLVISVNFNPDGS-----MLASTSAD 1551
QY 422 STVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHVS 481
DB 1552 KTVKLWRSHDGHLLHTFSGHNNVYSSFSFDPGRYIASASEDKTVKIWDIGHLLTLPQ 1611
QY 482 GTGGIFEVCNNAAGDKVGASASD 504
DB 1612 HQAGVMSAIFSPDGTLLISGSLD 1634

RESULT 6
AI2493

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2493
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1189 <KUR>
A;Cross-references: GB:BA000020; PIDN:BA078213.1; PID:gl7135667; GSPDB:GN00180
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7129
A;Genome: plasmid

Query Match
Best Local Similarity 16.1%; Score 439.5; DB 2; Length 1189;
Matches 109; Conservative 65; Mismatches 141; Indels 61; Gaps 9;

QY 167 GHESEVFIKAWNFVSDLLASGSDSTARIWLSNLSGSGTQLVLRHCIREGGQ-- 220
DB 686 GHDACVMSVVPHPVQILATAGEDNTIKLWELQSGCCLTKLQGHQHWKTIAPNSGGRIL 745
QY 221 -----DVPSNK-----DVTSLDNNSGTTLLATGSDYDGFARIW-TKDN 257
DB 746 ASGSPDQNVKLDIHTGKCVMTLQGHGTVTVSVAFNPKDNLNLSGSDSVQSVQSWDRKTR 805
QY 258 LASTLQGHQKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEAKQFPFHSAPALDV--D 315
DB 806 CLDTLKKHNRWSVAFHPQGHFLVSGGDDHAAKIWELGTGQCIKTFQCHSNATYIAH 865
QY 316 WOSNNTFASCSTDMCIHVCKLG-----ODRIKTFQHTNEVNAIKWDPTGNLLAS 366
DB 866 WE-HSLLASGHEDQTIKLDNLNLSHPKSNVNTNTHFRILQGHNSRVSVSVFSTGQLLAS 924
QY 367 CSDDMTLKIWSMKQNCVHDLQOHNKIYTIKWSPTGPTNNPNANMLASASFDSTVRL 426
DB 925 GSADRTIKLWSHPTGQCLHTLHGSGSWVAIAFSLD-----DKLLASGSYDHTVKI 975
QY 427 WVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTGALVHVSRTGGI 486
DB 976 WDVSSGQCLQTLQGHGGSVLAFAVSCDGTLPSSGVEKLVKQWDVETGYCLQT----- 1028
QY 487 FEVCNNAAGDKVGASA 502

Db 1281 DQQT-----IVSAGADNTVKLSRNTLL-----TTLEGHNEAVVQVIFSPD 1322
QY 181 SLLASGSGDSTARLWNLSENSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Db 1323 GRLIATAGADKTIITLWSRGN-----IL-----GTFAGHNEVNSLSFSPGNIL 1367
QY 241 ATGSDYGFARITWTDGNLSTLQGHKGFIPALKWNNKGNFILSAGVDKTTIWDIA-----295
Db 1368 ASGSDNTVRLWTVNRTLPKTFYGHKGSVSVRESNDGKITSLSLSTSTWKTWSLQKLL 1427
QY 296 -----HT-----GEAKQOPPHSAPALDWDQSN- 319
Db 1428 QTLASPLDVTSISFTPDNKNIVALASPDHTHLVNRQGLRLSLPGHNHWITSLSFSPNK 1487
QY 320 NTFASCSTDMCIHVCKLQGDPIKTFQCHTNEVNAIKWDPTGNLLASCSDMTLKIWSMK 379
Db 1488 QILASGSADKTIKLSV-NGRLLLKTLHGNGWVTDIKFSADGKNIVSASADKTIKWSL- 1545
QY 380 QDNCHVDLQCHNKSIYIKSPGTPGTPGNPNANLMLASAFDSTVRLWDVDRGICHTLT 439
Db 1546 DGRILTLQGHSAVWSVNLSPDQG-----TLASTSQDSTIKLNNLN-CELLYTLR 1595
QY 440 KHQEPVSVAFSPDGRVLASGSPKCVHIWNTQTGALVHSYRG-TGGIPEVCWNAAGDKV 498
Db 1596 GHSVDVNLSPSPDKTIASASDDGTIKLWVNPNGTLLKTFQHRGVRVSFSPDGKIL 1655
QY 499 GASASDGSVCVLDL 512
Db 1656 ASGGHDTTVKWNL 1669

RESULT 10
AC1889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:gl7130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 15.1%; Score 412.5; DB 2; Length 934;
Best Local Similarity 23.7%; Pred. No. 2e-21;
Matches 131; Conservative 113; Mismatches 196; Indels 113; Gaps 21;
QY 8 VNFVYRYLQESGSHAFYGIKSHISQSNINGALVPPAALISITQKGLQVYAEVSIN 67
Db 231 ISSLVTLIALGAGVAVLQGGKARMSYTK-----AISSEAFINANLEF- 277
QY 68 EDGLTFD---GRPIESLSDA-VMPDVQVQTOQAY-----RDKLAAQQAQAAAAA 118
Db 278 -DGLIASIRARRIKGTGIDANTRTQITETLQGSINFVRKNRLAEHDMLESVSFSD 336
QY 119 SQQSASKNNGENTANGENGATIANHNTDMVEVDGVEIPPNKAVVLRGHESEVP-ICAW 177
Db 337 SKFTATASRDKTVK-----IWSJDKKKQL-----VVRKXGEGFNVSF 376
QY 178 NPVSDLLASGSDSTARLWNLSENSTSGSTQLVLRHCIREGGQ--DVPSNKD-VTSLDWN 234
Db 377 SPDGTLMATGSDWNTAKIWS-----REGKRLHTLDGHEAVLEAFVS 418
QY 235 SEGTLTATGSDGPARITWTDGNLSTLQGHKGFIPALKWNNKGNFILSAGVDKTTIWD 294

Db 419 PDSQLLATASWDNTVKLWSREGKLLHTLEGHKKNVSIIFSPPQQLIATVGVWNTWKLWN 478
QY 295 AHTGEAKQOPPHSAPALDWDQ-----SNNTFASCSTDMCIHVCKLQGDPIKTFQCHT 349
Db 479 L-DGKELRTERFQHQ-----DMIVSVSPDQKQIATASGDRTVKLWSL-DGKELQTLRGHQ 532
QY 350 NEVNAIKWDPTGNLLASCSDMTLKIWSMKQDNCVHDLOCHNKSIYIKSPGTPG-----405
Db 533 NGVNSVTFSPDGKLIATASGDRTVKLWNSKQOB-LETLYCHTDAVNSVAFSPDGTSIATA 591
QY 406 -----TNNPNANLM-----LASASPDSTVRLWDV--DRGICI 435
Db 592 GNDKTAIKLWNLSPNSIIVRGHEDEVDFLVFSPNGKYIATASWDKTAKLWSIVGDKLOEL 651
QY 436 HTLKHOEPVSVAFSPDGRVLASGSPKCVHIWNTQTGALVHSYRG-TGGIPEVCWNA 494
Db 652 RFPNGHQVRVNSFSPDGKIIATTSWDKTAKLWNL-DGLQKTLIGHKQDVTWVNSVSPD 710
QY 495 GDKVGSASDGSV 507
Db 711 QOLIATASEDKTV 723

RESULT 11
AC1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:gl7135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283

Query Match 15.1%; Score 411.5; DB 2; Length 1747;
Best Local Similarity 27.4%; Pred. No. 5.5e-21;
Matches 114; Conservative 67; Mismatches 156; Indels 79; Gaps 10;
QY 165 LRHSESVFTICANVPVSDLLASGSGDSTARIWNL-----SENSTSGSTQ 208
Db 1310 LTGHERITSVKFSKILASASGDKTIKFWNTDGKFLKTAHAHQVNSINFSSDSKT 1369
QY 209 LV-----LRHCIREGGQV-----PSNK-----226
Db 1370 LVSGADSTMKVWKIDGTLLIKTISGRGEQIRDVTFSPDNKVIASASDXTVIRQLNYOK 1429
QY 227 ----DVTSLDWNSEGTLLATGSDYGFARITWTDGNLA-----STLQGHKGFIPALKWNNKKG 278
Db 1430 SQKSNVNSVFNPDGKTFASAGWDGNITWQRE-TLAHSLSTSIQKNQIITVSYSPDG 1488
QY 279 NFILSAGVDKTTIWDIAHTGEAKQOPPHSAPALDWDQ--SNNTFASCSTDMCIHVCKLG 337
Db 1489 KTIATASADNTIKLWDSQTLQKLTIKGHKDRITTLSPHPDNTIAGSGADKTIKWRVN 1548
QY 338 QDRPKTFQCHTNEVNAIKWDPTGNLLASCSDMTLKIWSMKQDNCVHDLOCHNKSIYI 397
Db 1549 DGQLRLTLTGHNDEVSFVNSFSPDQGLASGSDTNTVKIW-QTDGRLIKNITGHGLAIAVS 1607
QY 398 KWSPTGCTNNPNANLMLASAFDSTVRLWDVDRGICHTLTQHOEPVSVAFSPDGRYL 457
Db 1608 KFSPD-----SHTLASASWDNTIKLQVTDGKLIINNNGHIDGVTLSFSPDGEIL 1658

QY 458 ASGSFDCVHWNTOTGALVHSYRG-TGIFEVCHNAAGDKVGSASDGSVCVLDL 512
 Db 1659 ASGSADNTIKLWNPATLLKTLHGPKNITLAFSPDGKTLTLLSGEDAGVYVWNL 1714
 RESULT 12
 T40883
 WD repeat protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T40883
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
 A:Reference number: Z21954
 A:Accession: T40883
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-564 <WOO>
 A:Cross-references: EMBL:AL031764; PIDN:CAA21113.1; GSPDB:GN00068; SPDB:SPCC1235.09
 A:Experimental source: strain 972h-; cosmid c1235
 C:Genetics:
 A:Gene: SPDB:SPCC1235.09
 A:Map position: 3
 A:Introns: 18/1; 273/3; 413/3
 Query Match 14.3%; Score 390.5; DB 2; Length 564;
 Best Local Similarity 24.5%; Pred. No. 3.7e-20;
 Matches 133; Conservative 107; Mismatches 201; Indels 101; Gaps 24;
 QY 3 ISSDEVNLYRYLOESGFHSAPTFGIKSHIQSNIN---GALVPPAALISIIKGLQY 59
 Db 1 MDTNQVNYIIWYIKGCGYSHTKFAFERETGI--QNLDKQMGTCQVAGALVEILQKGLQY 58
 QY 60 VEAEV-----SINEDG--TLFQGRPIESLID-----AVMPDVVQT----- 94
 Db 59 VELEKHVYDHNSSNEEASKSIDG-----SLVNEPCKLPFYLTVPHCITTLTKADST 113
 QY 95 -----RQAYRDKLAQQOQAAAAAASQGSAGKNG-----ENTANGEENGHTIA 142
 Db 114 NGFCEHNNSNDHQLKILQDKSGSPSPVMPFKDKIEKRDITWADSNVEKDPARPIA 173
 QY 143 NNHTDMVEVDGVEIIPNKAVLRLGHE--SEVF-----ICA-WNPVSD-----LLA 185
 Db 174 VYNSSPV-----TEITEIKQVTTGGEDIKSDPFKVPKHPVTCDMPRLQENYHYVE 228
 QY 186 SGSGDSTARLWNL-----ENSTSGTQLVLRHCIRGGQDVPSNKDVTSLDNSEGTLLA 241
 Db 229 FSIQMTNATLASVSIQEQNDPRAKTD---YCL-----QSSFDNQDITGVANNNGSFLA 280
 QY 242 TGSYDGFARTWDXGNLASTLGHKGPFPALKKWKGNFILSAGVDKTTIINDAHTGEAK 301
 Db 281 YAFSGVIEIYDSHGSGQILSFHNKGPVLSLWSGTDITYLAAGSADGTTILFD---QLK 336
 QY 302 Q-QPPEH--SAPALDWDWOSNNTFASCSTDMCIHVCKLGQDRPIKTF-QGHNEVNAIKW 357
 Db 337 QTCYSIDTLASSVLDIEMISFDEFVTSVGEGLRVYKVDGKAPVSTVSHAHNSIVALRY 396
 QY 358 DPTGNLLASCDMDTLKISMKQD---NCVHDLQHNKEIYTIKWSPTGPGNPNANLM 414
 Db 397 NLRISLLLTASSDTTVKLSRSGDAGAFGLH-VFSSFPVNCIDW-----NLREGTPI 448
 QY 415 LASASPSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRVLAGSPDKCVHKNWTTOTG 474
 Db 449 LAVAS-NSIVSMYNALSLQQLAVFMHTAPVLSLSFHNRYLATGDTSGGVCIVSCKTA 507
 QY 475 AL 476
 Db 508 XL 509

RESULT 13

S76086
 beta transducin-like protein, 190K - Synechocystis sp. (strain PCC 6803)

N;Alternate names: protein sil0163
 C;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
 C;Accession: S76086
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76086
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1693 <KAN>
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:G1001396; PIDN:BA010064.1; PID:d1010711
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:1051-1084/Domain: WD repeat homology <WD01>
 F:1092-1125/Domain: WD repeat homology <WD02>
 F:1133-1166/Domain: WD repeat homology <WD03>
 F:1174-1207/Domain: WD repeat homology <WD04>
 F:1256-1289/Domain: WD repeat homology <WD05>
 F:1297-1330/Domain: WD repeat homology <WD06>
 F:1338-1371/Domain: WD repeat homology <WD07>
 F:1420-1453/Domain: WD repeat homology <WD08>
 F:1461-1494/Domain: WD repeat homology <WD09>
 F:1502-1535/Domain: WD repeat homology <WD10>
 F:1584-1617/Domain: WD repeat homology <WD11>
 F:1625-1658/Domain: WD repeat homology <WD12>
 Query Match 14.3%; Score 388.5; DB 2; Length 1693;
 Best Local Similarity 28.5%; Pred. No. 2.3e-19;
 Matches 107; Conservative 52; Mismatches 138; Indels 79; Gaps 10;
 QY 100 RDLAQOQAAAAAASQGSAGKNGENTANGE--ENGAHTIANNHTDMVEYDGDVEI 157
 Db 1362 RDKTARLWTTTEGCVAVLADHOGVWREGQFSPDQWIVTGS---ADKTAQLWNLV-- 1414
 QY 158 PPNKAVLRLGHESEVFICANNPVSDDLASGSDSTAIRNLSENSTSGTQLVLRHCIRE 217
 Db 1415 --KKLTVLRLGHQDVLNVRFPDSQYIVTASKDGTAFAVWN-----NTGRELAVLRH-- 1463
 QY 218 GGQDVPSNKDVTSLDNSEGTLLATGSDYGFARITWTKDGNLSTLGHKGPFPALKWKK 277
 Db 1464 -----YEKNIFAEFPADGQFIVTASDDNTAGIWEIVGREVCICRGEHGFVFAQFSAD 1517
 QY 278 GNFLSAGVDKTTIINDAHTGEAKQPPFHSAPALDWDWOSNNTFASCSTDMCIHVCKLG 337
 Db 1518 SRYLTASVDNTARIDF-----LG 1537
 QY 338 QDRPIKTFQGHNEVNAIKWDPITGNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTI 397
 Db 1538 --RPLLLTAGHQSIVYQARFSPGEGNLIATVSADHTARLWD-RSGKTVAVLYGHQGLVGT 1594
 QY 398 KWSPTGCTGNPNANMLASAFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYL 457
 Db 1595 DWSPDGQ-----MLVTASNDGTARLWDLIS-GRELLTLEGHGNWVRSAAEFPDGRW 1644
 QY 458 ASGSFDCVHWNTOT 473
 Db 1645 LTSSADGTAKLWPKVT 1660
 RESULT 14
 AH2195
 hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AH2195
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2195
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-676 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA074818.1; PID:gl7132214; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3119

Query Match 14.0%; Score 382.5; DB 2; Length 676;
Best Local Similarity 23.6%; Pred. No. 1.8e-19;
Matches 123; Conservative 85; Mismatches 195; Indels 119; Gaps 12;
QY 21 FHSATFTGKSHISOSNGALVPPAALISITOKGLQVVEA-----EVS 65
DB 226 FSLGATCFHLLTGTFNSL-----FVEQGSVSVESQOQYWNITSNDREGEYL 273
QY 66 INEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRKLAQQQAAAA-----AAAAASQOG 122
DB 274 VKVNLKLETDIORRYQSADEVVNDLIKQSLLSRLKTTIPKSAIFRSWSASTSLTAST 333
QY 123 SAKNGENTANGENGAAHTIANNHTDMVEVDG-----DVEIPPENKAVVLRG 167
DB 334 TKQAKWKLNGRLK-QOLLNNTMSALLGVGVGHLSQLPLQITKFSFISTQP----YTLKG 389
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKD 227
DB 390 HASDVNSVAFSPNGEFLASGSDDKTIKVMNL-----420
QY 228 VTSLDWNSEGTLLATGSDYDFARIWTKDGNLSTLQGHKGPFIKXNKKGNFILSAGVD 287
DB 421 -----KNKOKIHTLPGHSGWVAIAFSPDGKTLASTGAD 454
QY 288 KTTIIDAHTGEAKQPPHSPALDYDWO-SNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 455 KTIKLNWLTATGEIRHLKGHSQGVASVAFSPDGKTLASGLDKTIKLNPNATGKEIRTLQ 514
QY 347 GHTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPT 406
DB 515 EHSQGVANVAFSPDGKTLASGSDKTIKLNLTTSKVIHTLKGHSDLVMSVAFN-----568
QY 407 NNPANMLASASFDSTVRLWDVDRGICHTLKHQEPVYVAFSP-DGRYLASGSFDCX 465
DB 569 ----SDSQTLAGSKDKTIKLNWLTSTGKTIPTLRGHSQDKNSVAYVPEDSTVLASGSNDNT 625
QY 466 VHIWNTOTGALVHSY-RGTGCIFFVCWNAAGDKV GASADGS 506
DB 626 IKLWNLTTGEIIRTKRDSGYIYSIVISPDGRNL---ASGGS 664

RESULT 15
AB2410
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2410
R;kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2410
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1551 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA076533.1; PID:gl7133971; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4834

Query Match 14.0%; Score 382; DB 2; Length 1551;
Best Local Similarity 22.6%; Pred. No. 6.1e-19;
Matches 122; Conservative 94; Mismatches 193; Indels 130; Gaps 18;
QY 48 ALISIIQKGLQYVEAEYSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRKLAQQQ 107
DB 903 ALMSAMRSG-----KALQALVKGDSLAKYPATSPLLALQITLDNIQERNQ-----FQGHQ 953
QY 108 AAAAAAASQGSARNGENTANGENGAAHTIANNHTDMVEVDGVEIPPENKAVVLRG 167
DB 954 AWRSVSFSDRGQVILTASDDCTAR-----LWNLQG-----KQLISLQG 992
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLV-----RHCIR 216
DB 993 HEDTIWISANFSPDGKTIATASDSTRNLNFSGOOLAKFQGHQGVRSVSFSPDGKHIAT 1052
QY 217 EG-----GQDV-----PSNKDV-TSLD-----WNSGTL-- 240
DB 1053 AGDDHTAELWFSGQQLVQFPGHQGYVWCISFSPDGKHIATAADRIVRLWNLKGLVR 1112
QY 241 -----ATGSYDGPARIWTKDGNLSTLQGHKGPFIKXNKKGNF 280
DB 1113 FPGHQCVCVWVSFSPDSQYIATASDSTRNLNFSGOOLAKFQGHQGVRSVSFSPNGQY 1172
QY 281 ILSAGVDKTTIIDAHTGEAKQPPHSPALDYDWOQNNNTF-ASCSTDMCIHVCKLGQD 339
DB 1173 IATSSDRTARWNLN-QQLAQFSGHODYVRSVSFSPDGKTIATASDRIVRLWHLNKQ 1231
QY 340 RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIK 399
DB 1232 Q-FSAFQGHQSTVRSVDFSPDGQKVTAAADRTVRLWNLKGEELIQFL-GHRGKVSVSF 1289
QY 400 SPTGPTNNPNANMLASASFDSTVRLWDVDRGICHTLKHQEPVYVAFSPDGRYLAS 459
DB 1290 SPDGK-----VIATTSDDRTVRLWDI-TGQLLOQFPQGHQGVRSVSFSPDGKHIAT 1339
QY 460 GSFDCVCHIMNTQTGALV-----HSYRGTGCIFFVCWNAAGDKV 499
DB 1340 ASSDLTLRLSLDGGELMQFKGHDKVVYVFSFCNGQHIATAADCTARLWNLAGRQVG 1398

Search completed: August 9, 2004, 16:48:48
Job time : 15.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-2
Perfect score: 2726
Sequence: 1 MSISDEVNFLVRYLQESG.....GRKVGASASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2691	98.7	514	11	Q8BHJ5	Q8bhj5 mus musculus
2	2688	98.6	514	11	Q8CBQ4	Q8cbq4 mus musculus
3	2687	98.6	514	11	Q9EOD4	Q9eqd4 mus musculus
4	2644.5	97.0	519	13	Q7SZM9	Q7szm9 xenopus lae
5	2484.5	91.1	527	11	Q8BYQ4	Q8byd4 mus musculus
6	2480.5	91.0	527	11	Q8BMM0	Q8bmm0 mus musculus
7	2437	89.4	577	4	Q8EUY2	Q8euy2 homo sapien
8	1972.5	72.4	412	11	Q8COA1	Q8co01 mus musculus
9	1904.5	69.9	700	5	Q9XZK1	Q9xzk1 drosophila
10	1387	50.9	613	10	Q9FN19	Q9fn19 arabidopsis
11	1096	40.2	201	11	Q8VEG3	Q8veg3 mus musculus
12	923.5	33.9	524	5	Q9SRU9	Q9srj9 drosophila
13	477	17.5	1356	3	Q8XLP4	Q8xlp4 drosophila
14	474	17.4	1356	3	Q8XLP5	Q8xlp5 podospora a
15	472	17.3	1356	3	Q8XLP3	Q8xlp3 podospora a
16	449	16.5	1376	3	Q8XLP2	Q8xlp2 podospora a

17	439.5	16.1	1189	16	Q8YL09	Q8yl09 anabaena sp
18	439	16.1	1227	16	Q8ZOR1	Q8zor1 anabaena sp
19	437	16.0	1708	16	Q8Y212	Q8y212 anabaena sp
20	414	15.2	1711	16	Q8Z019	Q8z019 anabaena sp
21	412.5	15.1	934	16	Q8Y223	Q8y223 anabaena sp
22	411.5	15.1	1747	16	Q8Z020	Q8z020 anabaena sp
23	390.5	14.3	564	3	O74845	O74845 schizosacch
24	382.5	14.0	676	16	Q8YSG6	Q8ysg6 anabaena sp
25	382	14.0	1551	16	Q8YMU3	Q8ymu3 anabaena sp
26	376	13.8	415	4	Q8N136	Q8n136 homo sapien
27	372	13.6	1233	17	Q8TWX4	Q8tmx4 methanosarc
28	371.5	13.6	502	3	O74855	O74855 schizosacch
29	370.5	13.6	481	5	Q9VPR4	Q9vpr4 drosophila
30	370.5	13.6	488	5	Q8T4A2	Q8t4a2 drosophila
31	370	13.6	415	4	Q8N776	Q8n776 homo sapien
32	368.5	13.5	304	11	Q9D4T2	Q9d4t2 mus musculu
33	368.5	13.5	304	16	Q8Z054	Q8z054 anabaena sp
34	368.5	13.5	1189	16	Q8YTD1	Q8ytd1 anabaena sp
35	368	13.5	411	5	O96698	O96698 drosophila
36	364	13.4	1241	2	Q9XBD8	Q9xbd8 amycolatops
37	362.5	13.3	480	5	O96995	O96995 drosophila
38	361.5	13.3	410	13	Q803D2	Q803d2 brachydanio
39	361	13.2	476	13	O93531	O93531 xenopus lae
40	361	13.2	476	13	Q7ZXK9	Q7zxk9 xenopus lae
41	360	13.2	265	16	Q8Y216	Q8y216 anabaena sp
42	359	13.2	352	10	O80990	O80990 arabidopsis
43	358.5	13.2	1430	16	Q98HK1	Q98hk1 rhizobium l
44	355.5	13.0	410	13	Q90ZL4	Q90z14 xenopus lae
45	354	13.0	339	11	Q9DCZ7	Q9dcz7 mus musculu

ALIGNMENTS

RESULT 1

Q8BHJ5 ID Q8BHJ5 PRELIMINARY; PRT; 514 AA.
AC O8BHJ5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE IRA1 protein.
GN IRA1 OR 803049H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002)."
RL EMBL; AK029595; BAC26526.1; -
DR ENBL; AK033347; BAC28241.1; -
DR PIR; PT0651; PT0651
DR MGI; MGI:2441730; Iral
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PFO0400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS00936; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00682; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 98.7%; Score 2691; DB 11; Length 514;
 Best Local Similarity 98.6%; Pred. No. 6.2e-179;
 Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSDEVNFVYRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNFVYRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOAAAAAATNQ 120
 QY 121 QGSAGKNGTANGENGAGHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVFIKAWNPV 180
 DB 121 QGSAGKNGTANGENGAGHTIANNHTDMVEVDGVEIPSNKAVILRGHSEVFIKAWNPV 180
 QY 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 DB 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 QY 241 ATGSYDGFARITWKDGNLSTLQGHKGPITFALKWKKGNFILSAGVDKTTIWDHAHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTLQGHKGPITFALKWKKGNFILSAGVDKTTIWDHAHTGEA 300
 QY 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480
 DB 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480
 QY 481 RGTGGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514

RESULT 2
 ID Q8CEG4 PRELIMINARY; PRT; 514 AA.
 AC Q8CEG4;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; Pubmed=12466851;
 RA The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK036064; BAC29294.1; -.
 DR PIR; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.

DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 514 AA; 55689 MW; 13BED3753A725029 CRC64;

Query Match 98.6%; Score 2688; DB 11; Length 514;
 Best Local Similarity 98.4%; Pred. No. 1e-178;
 Matches 506; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISSDEVNFVYRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNFVYRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFQGRPIESLSLIDAVMPDVQTRQAYRDKLAQOAAAAAATNQ 120
 QY 121 QGSAGKNGTANGENGAGHTIANNHTDMVEVDGVEIPPNKAVILRGHSEVFIKAWNPV 180
 DB 121 QGSAGKNGTANGENGAGHTIANNHTDMVEVDGVEIPSNKAVILRGHSEVFIKAWNPV 180
 QY 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 DB 181 SDLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLL 240
 QY 241 ATGSYDGFARITWKDGNLSTLQGHKGPITFALKWKKGNFILSAGVDKTTIWDHAHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLSTLQGHKGPITFALKWKKGNFILSAGVDKTTIWDHAHTGEA 300
 QY 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQOPPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480
 DB 421 DSTVRLMDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480
 QY 481 RGTGGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCNNAAGDKVGASASDGSVCVLDLRK 514

RESULT 3
 ID Q9EQD4 PRELIMINARY; PRT; 514 AA.
 AC Q9EQD4;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Zhang X., Dormay S., Basch R.;
 RT "Identification of four human cDNAs that are differentially expressed
 RT by early hematopoietic progenitors."
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268195; AAG44738.1; -.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.

DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; LISH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 514 AA; 55689 MW; 6A72CE68A40C141F CRC64;

Query Match
Best Local Similarity 98.6%; Score 2687; DB 11; Length 514;
Matches 506; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSISSEVNFVLYRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISSEVNFVLYRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 120
Db 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 120
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 120
Db 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 120
QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPNKAVLGRHSEVFI 180
Db 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPNKAVLGRHSEVFI 180
QY 181 SLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSKNDVTSLDWNS 240
Db 181 SLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSKNDVTSLDWNS 240
QY 241 ATGSYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKNGFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKNGFILSAGVDKTTIIWDAHTGEA 300
QY 301 KQGFPHSAPALVDWQSNNTFASCTDWCIVHCKLGQDRPIKTFQGHNEVNAI 360
Db 301 KQGFPHSAPALVDWQSNNTFASCTDWCIVHCKLGQDRPIKTFQGHNEVNAI 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKKEIYTIKWSPTGPTNNPNANMLAS 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKKEIYTIKWSPTGPTNNPNANMLAS 420
QY 421 DSTVRLWDVDRGICHTLTIKHQPYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 480
Db 421 DSTVRLWDVDRGICHTLTIKHQPYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 480
QY 481 RGTGIFFCVWNAAGDKVGASDGSVCVLDLRK 514
Db 481 RGTGIFFCVWNAAGDKVGASDGSVCVLDLRK 514

Query Match
Best Local Similarity 97.0%; Score 2644.5; DB 13; Length 519;
Matches 500; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 MSISSEVNFVLYRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISSEVNFVLYRYLOESGFHSATFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 118
Db 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOAAAAAASASO 120
QY 119 --SOQSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPNKAVLGRHSEVFI 175
Db 121 TPNQQQPAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPNKAVLGRHSEVFI 180
QY 176 AWPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSKNDVTSLDWNS 235
Db 181 AWPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSKNDVTSLDWNS 240
QY 236 EGTLLATGSDYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKNGFILSAGVDKTTIIWDA 295
Db 241 EGTLLATGSDYDGFARIWTKDGNLASTLGQHKGPFIKWKNGKNGFILSAGVDKTTIIWDA 300
QY 296 HTGEAKQGFPHSAPALVDWQSNNTFASCTDWCIVHCKLGQDRPIKTFQGHNEVNAI 355
Db 301 HTGEAKQGFPHSAPALVDWQSNNTFASCTDWCIVHCKLGQDRPIKTFQGHNEVNAI 360
QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKKEIYTIKWSPTGPTNNPNANML 415
Db 361 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLOAHNKKEIYTIKWSPTGPTNNPNANML 420
QY 416 ASASFDSTVRLWDVDRGICHTLTIKHQPYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 475
Db 421 ASASFDSTVRLWDVDRGICHTLTIKHQPYSVAFSPDGRYLASGSPDKCVHIWNTQTGA 480
QY 476 LVHSYRTGTGIFFCVWNAAGDKVGASDGSVCVLDLRK 514
Db 481 LVHSYRTGTGIFFCVWNAAGDKVGASDGSVCVLDLRK 519

RESULT 5
Q8BYQ4 PRELIMINARY; PRT; 527 AA.

ID Q8BYQ4
AC Q8BYQ4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transducin (Tbalx protein).
GN TBLX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs";
RN Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalko U., Schmutz D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AK038674; BAC30092.1; -
DR ENBL; BC043105; AH43105.1; -
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tblx.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;

Query Match 91.1%; Score 2484.5; DB 11; Length 527;
Best Local Similarity 87.9%; Pred. No. 1.5e-164;
Matches 463; Conservative 28; Mismatches 21; Indels 15; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSITSDEVNLFVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQQAAYDKLAQOQAAAAA----- 115
DB 61 EAEISINEDGTVFDGRPIESLSLIDAVMPDVVQTRQQAAYDKLAQOQAAAAA----- 120
QY 116 -----AAASQGSAGKNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVALR 166
DB 121 ATSTAATPAAAAQONPKNGEATVNGEANGAHAI -NNHSPKMEIDGVEIPPSKATVLR 179
QY 167 GHSEVFICAWNPVSDLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNK 226
DB 180 GHSEVFICAWNPVSDLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGHVDPSNK 239
QY 227 DVTSLDWNSGTLTATGSGYDGFARITWKGDLASTLGQHGKPIFALKWKNKGNFILSAGV 286
DB 240 DVTSLDWNSDGTLLATGSGYDGFARITWKGDLASTLGQHGKPIFALKWKNKGNFILSAGV 299
QY 287 DKTTIIWAHTGEAKQOPPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIIWAHTGEAKQOPPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPVKTFQ 359
QY 347 GHTNEVNAIKWDPTGNLLACSDDMTLKIWSMKQDNVCHDLQOHKEIYTIKWSPTGPGT 406
DB 360 GHTNEVNAIKWDPSGMLLASCSDDMTLKIWSMKQDACVHDLQAHSEIYTIKWSPTGPGT 419
QY 407 NNPANLKLASASPDSTVRLWDVDRGICHTLTKEHPVYSVAFSPDGRYLASGSPDKCV 466

```



```

QY 347 GHTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHVDLQHNKEIYTIKWSPTGCT 406
Db 360 GHTNEVNAIKWDPGSMGLASCDDMTLKIWSMKQACVHDLOAHKSEIYTIKWSPTGCT 419
QY 407 NNPANMLASAFSTSVRLWVDVGRICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCV 466
Db 420 SNPNINMLASAFSTSVRLWVDVGRICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCV 479
QY 467 HWTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 514
Db 480 HIWNTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 527

RESULT 7
Q86UY2 PRELIMINARY; PRT; 577 AA.
AC Q86UY2;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE TBLIX protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Ioshizuka S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBAJ databases.
DR EMBL: BC052304; AAHS2304.1; -.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001880; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

```

```

Query Match      89.4%; Score 2437; DB 4; Length 577;
Best Local Similarity 86.0%; Pred. No. 3.4e-161;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

```

```

QY 1 MSISSDEVFLVRYLOESGFSSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60
Db 52 MSITSDEVFLVRYLOESGFSSAFTFGIESHISQSNINGTLVPPAALISITQKGLQYV 111
QY 61 EAEYSINEDGTLDPGRPIESLSLIDAVMPDVQVOTRQQAQKLAQQAQAAAAA--- 117
Db 112 EAEISINEDGTVFGRPIESLSLIDAVMPDVQVOTRQQAQKLAQQAQAAAAAATAA 171
QY 118 -----ASQGSAGKNGENTANGENGAHTIANNHTDMMEVDGVEIPPNKAVYLRG 167
Db 172 ATAATTTAGVSHONPSKNEATVNGENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 230
QY 168 HSEVEFCANPNVSDLLASGSDSTARINWLSNSISGSTOLVLRHCIRGGQDVPNSKD 227
Db 231 HSEVEFCANPNVSDLLASGSDSTARINWLSNSISGSTOLVLRHCIRGGQDVPNSKD 290
QY 228 VTSLDNMSGTLLATGSDYGFARIWTKDGNLSTLGHKGPIFALKWNKNGNFTLSAGVD 287
Db 291 VTSLDNMSGTLLATGSDYGFARIWTKDGNLSTLGHKGPIFALKWNKNGNFTLSAGVD 350
QY 288 KTIINDAHTGEAKQPPHSAFALVDVQSNNTFASCSTDHICVHCKLQODRPIKTFQG 347
Db 351 KTIINDAHTGEAKQPPHSAFALVDVQSNNTFASCSTDHICVHCKLQODRPIKTFQG 410
QY 348 HTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHVDLQHNKEIYTIKWSPTGCTN 407
Db 411 HTNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCHVDLQHNKEIYTIKWSPTGCTN 470
QY 408 NPNANMLASAFSTSVRLWVDVGRICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 467
Db 471 NPNANMLASAFSTSVRLWVDVGRICHTLTKHQEPVYSVAFSPDGRYLAGSFDKCVH 530
QY 468 IWTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 514
Db 531 IWTQSGLVHSYRGTTGIFEVCNNAAGDKVGASDGSVCVLDLRK 577

```

RESULT 8

```

Q8C0A1 PRELIMINARY; PRT; 412 AA.
AC Q8C0A1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Transducin (Fragment).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK031937; BAC27612.1; -.
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR001880; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
FT NON TER 1
SQ SEQUENCE 412 AA; 44356 MW; EB78910E6D9E5237 CRC64;

```

Query Match 72.4%; Score 1972.5; DB 11; Length 412;
 Best Local Similarity 87.6%; Pred. No. 4.4e-129;
 Matches 360; Conservative 22; Mismatches 28; Indels 1; Gaps 1;

QY 104 AQQQAAAAAASQGSAGKNGTANGENGGAHTIANDHMDMEVDGVEIPPKAV 163
 DB 3 AATATSTAAATPAAAAQQPPKNGEATVNGEENGAAHAI-NHRSKPMEDGDEVEIPPSKAT 61
 QY 164 VLRGHESEVFCAMPVPSDLLASGGSDSTARVWLSNSTSGTQVLRHCIRREGGQDVP 223
 DB 62 VLRGHESEVFCAMPVPSDLLASGGSDSTARVWLSNSTSGTQVLRHCIRREGGQDVP 121
 QY 224 SNKDVTSLDWNBEGTLLATGSDYGFARVWKDGNLSTLQGHKQIPFALKNNKGNFILS 283
 DB 122 SNKDVTSLDWNBEGTLLATGSDYGFARVWKDGNLSTLQGHKQIPFALKNNKGNFILS 181
 QY 284 AGVDKTTIWDHTGAKQFPFHSAPALDQVQSNTPFASCTDMCHVCKLGQDRPK 343
 DB 182 AGVDKTTIWDHTGAKQFPFHSAPALDQVQSNTPFASCTDMCHVCKLGQDRPK 241
 QY 344 TQCGHTNEVNAIKWDPTEGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTG 403
 DB 242 TQCGHTNEVNAIKWDPTEGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTG 301
 QY 404 PGTNPNANMLASAFDSFTVRLWDVDRGICHTLTQCEPVYSVAFSPDGRYLASGSFD 463
 DB 302 PATSNFNSNIMLASAFDSFTVRLWDVDRGICHTLTQCEPVYSVAFSPDGRYLASGSFD 361
 QY 464 KCVHIWNTQSGLVHSYRGTTGGIFECVWNAAGDKVGASDGSVCVLDLRK 514
 DB 362 KCVHIWNTQSGLVHSYRGTTGGIFECVWNAAGDKVGASDGSVCVLDLRK 412

RESULT 9

Q9XZK1 ID Q9XZK1 PRELIMINARY; PRT; 700 AA.
 AC Q9XZK1
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DE EBI protein.
 GN EBI OR C04063.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Berkely.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos E., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Durbin K., Doup L.E., Dowes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalushon K.A., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Renington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidon-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Stradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN SEQUENCE FROM N.A.
 RP MEDLINE=9234084; PubMed=10215623;
 RX Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
 RA "ebi regulates epidermal growth factor receptor signaling pathways in
 RT Drosophila.";
 RL Genes Dev. 13:954-965 (1999).
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Berkely.
 RA Harris N.L., Agbayani A., Arcaina T.T., Baxter E., Blazek R.G.,
 RA Chavez C., Chew M., Doyle C.M., Farfan D.E., Flanagan J.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim S.H., Lee B.,
 RA Lomotab M.A., Mak J., Mazda P., Moshrefi A.R., Moshrefi M., Nixon K.,
 RA Pacleb J.M., Park S., Pfeiffer B., Punch E., Snir E., Twomey B.,
 RA Wan K.H., Zhang R., Zieran L., Rubin G.M.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003589; AAF5101.1; -;
 DR EMBL; AF146345; AAD35017.1; -;
 DR FLYBase; FBgn0023444; ebi.
 DR GO; GO:0000074; P:regulation of cell cycle; IMP.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINERPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SMC0667; Lish; 1.
 DR SMART; SMC0320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00037; MYB_1; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 3.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79PB7 CRC64;

Query Match 69.9%; Score 1904.5; DB 5; Length 700;
 Best Local Similarity 54.8%; Pred. No. 4.9e-124;
 Matches 385; Conservative 51; Mismatches 75; Indels 191; Gaps 8;
 QY 1 MSISDSVNFVLYRLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60
 DB 1 MSFSDSVNFVLYRLOESGFLSHAYVFGIESHSQSNINGALVPPAALLTIQKGLYT 60
 QY 61 EAVYSINEDGTLFGRIEISLIDAVMDV-----VQTRQ----- 96
 DB 61 EVESVSGEDGEV--ARPIEGLSLIDAVMPVKPLIKTEPGKPGAVDSAPAGNQNN 118
 QY 97 -----QAYRDKLA-----QQQAAAAAASQGSAGKNGENT 130
 DB 119 NAKPEIKIEPTGTVAGSAGGNKIAGSTGTSTPDQASAEVDSGNAANNAGTYAGNNG 178
 QY 131 ANGBE----- 135

```

Db 179 AGNQASTGGSNSTSTPACGDLAAPGASQKQSNSEAGSSSGNAGNANATSTDDAASS 238
QY 136 ---NG-----AHTIANNHDTM-----148
Db 239 TSTNGNSSTSSVQPTSLGTAGTIVTSNPDAASGGASTATSKAPSGAVTIRGQAQ 298
QY 149 -----MEVDGVEIPPKNKAVVLRGHESEV 172
Db 299 GNVQSGSSNAOSSAPSGTISSTSGGAGTAAALVPMIDENIEIPESKARVLRGHESEV 358
QY 173 FICAMPVPDLASGSGDSTARIWNLSENSTSGTOLVLRHCIRGGQDVPNSKDVTSID 232
Db 359 FICAMPSPDLASGSGDSTARIWNSDANTH-SNQLVLRHCIRGGAEVPSNKDVTSID 417
QY 233 WNSEGTLATGSDYGFARITWTKGNLSTLGOKGPIFALKWKKGNFILSAGVDKTTII 292
Db 418 WNCDSGLLATGSDYGVARIWTKTDLGLASTLGQKGPFIKWKKNKCNVILSAGVDKTTII 477
QY 293 WDAHTGEAKQPPFHSAPALDWDQSNNTFASCTDMCHVCKLGQDRPIKTFQHTNEV 352
Db 478 WDASTGCTQCPAFHSAPALDWDQTNQAFASCTDQRIHVCRGLGVNBPITFKGHTNEV 537
QY 353 NAIKMDPTGNLLASCSDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPGTGPTGNPNAN 412
Db 538 NAIKWCPCQGLLASCSDMTLKIWSMNRDRCCHDLQHSKEIYTIKWSPGTGPTGNPNAN 597
QY 413 LMLASASPDSTVRLWDVDRGICIHILTKHQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472
Db 598 LILASASPDSTVRLWDVDRGICIHILTKRTEPVSVAFSPDGHKLASGDFKCVHINSTQ 657
QY 473 TGAHVHSYGTGGIPEVCNNAAGDKVGASASGSCVCLDLRK 514
Db 658 TQQLVHSYKGTGGIPEVCNNSKGTGASASGSCVFLDLRK 699

RESULT 10
Q9FN19 ID Q9FN19 PRELIMINARY; PRT; 613 AA.
AC Q9FN19;
DT 01-WAR-2001 (TReMBLrel. 16, Created)
DT 01-WAR-2001 (TReMBLrel. 16, Last sequence update)
DE Genomic DNA, chromosome 5, TAC clone:K8K14 (AT5g67320/K8K14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

```

RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057696; AAL15328.1; -
DR EMBL; AY143932; AAN28871.1; -
DR InterPro; IPR006594; LISH.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODom; PD000018; WD40; 2.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00882; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 613 AA; 69772 MW; 04F40772311C0E76 CRC64;

Query Match 50.9%; Score 1387; DB 10; Length 613;
Best Local Similarity 44.2%; Pred.No. 4.le-88;
Matches 272; Conservative 91; Mismatches 144; Indels 108; Gaps 7;

QY 2 SSSDEYNELVRYLQESGFSHSAFTGKISHISQSNINGALVPPALISIIOKGLOYVE 61
Db 3 SLTSVNLFLVRYLQESGFTHAFTLGYEAGINKSNIDGNMVPFGALIKFVQKGLGYME 62
QY 62 -----AEVSINEDGTLFDGRTIESLSLIDAVMPDVVQTRQQAYRDKLAQCCAAAAA 114
Db 63 MEANLSNSEVDIEDPSFF--QPLDLISKVDKQELQDMLREKKRDMEKERDRSKNDK 120
QY 115 AAAASQOQSAKNGENTANGE-----GDVE-----ENGAHTIAN 143
Db 121 GVEREHGDRNRAKEKDRHEKQKEREKEREKEREKEREKEREKEREKEREKEREIFE 180
QY 144 NHTDMMEVD-----IPPKNKAVVLRGHESEVFICANNPVSDLLASGSGDSTARINWLSNSTS 204
Db 181 REKDLKLEKEREIEREREKEREKEREKEREKEREKEREKEREKEREKEREKEREKERE 240
QY 157 -----TTPKNKAVVLRGHESEVFICANNPVSDLLASGSGDSTARINWLSNSTS 204
Db 241 DIVMTPTTSQTSHPNSDVRILEGHTSEVCACAWSPSASLLASGSGDATARIWISPEGSFK 300
QY 205 -----GSTOLVLRHCIRGGQDVPNSKDVTSLDWNSGTLTATGSDYDGFARITWTKGNL 258
Db 301 AVHTGRNINAILKHA---KGKSNESKDVTTLDWNGEGTLTATGSCDQGARITWTLNGEL 357
QY 259 ASTLGHQHGPIFALKWKKGNFILSAGVDKTTIIWDAHTGEAKQPPFHSAPALDWDQSN 318
Db 358 ISTLSKHKGPIFSLKWNKGDYLLTGSVDRTAVVMDVKABEWKQPFHFHSGPTLDVDWRN 417
QY 319 NNTFASCSDDMCIHVCKLGQDRPIKTFQHTNEVNAIKWMDPTGNLLASCSDMTLKIWSM 378
Db 418 NVGFATSTDSMIYLCIKIGETRPAAKFTTGHQGEVNCVVKWDPTGSLASCSDSTAKIWN 477
QY 379 KQNCVHDLQHNKEIYTIKWSPGTGPTGNPNANLMLASASFDSTVRLWDVDRGICIHIL 438
Db 478 KQSTFVHDLREHTKEIYTIKWSPTGPTGNPNKQLTLASASFDSTVKVLDABLKGKMLCSF 537
QY 439 TKHQEPVSVAFSPDGRYLASGDFKCVHIWNTQTGALVHSYGTGGIPEVCNNAAGDKV 498
Db 538 NGRHREPVSVAFSPNGEYIASGLSKDSIHSIKGKIVKTYKTGNGGIFEVCWNKGNKI 597
QY 499 GASASGSCVCLDLR 513

```

Db 598 AACFADNSVCVLDLR 612

RESULT 11

Q8VEG3 PRELIMINARY; PRT; 201 AA.

AC Q8VEG3; (1)

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Similar to IRAL protein (Fragment).

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR ENBL; BC018512; AAH18512.1; -

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 4.

DR ProDom; PD000018; WD40; 2.

DR SMART; SM00320; WD40; 4.

DR PROSITE; PS00678; WD_REPEATS_1; 2.

DR PROSITE; PS00682; WD_REPEATS_2; 3.

DR PROSITE; PS00294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.

FT NON TER

SQ SEQUENCE 201 AA; 22038 MW; 6B945F137B491818 CRC64;

Query Match 40.2%; Score 1096; DB 11; Length 201;

Best Local Similarity 99.0%; Pred. No. 1.5e-68;

Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 314 VDMQSNNTFASCTDCIHHVKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTL 373

Db 1 VDMQSNNTFASCTDCIHHVKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDDMTL 60

QY 374 KINSMKQDNCVHDLQHNKEIYTIKWSPTGPGTNNENALMLASAFDSTVRLWDVDRGI 433

Db 61 KINSMKQDNCVHDLQARNKEIYTIKWSPTGPGTNNENALMLASAFDSTVRLWDVDRGI 120

QY 434 CIHTLTKHQPVYSVAFSPDGRVLAGSPDKCVHIWNTQTGALVHSYRGTFGEVCWNA 493

Db 121 CIHTLTKHQPVYSVAFSPDGRVLAGSPDKCVHIWNTQTGALVHSYRGTFGEVCWNA 180

QY 494 AGDKVGASADSGVCVLDLRK 514

Db 181 AGDKVGASADSGVCVLDLRK 201

RESULT 12

Q9SRJ9 PRELIMINARY; PRT; 524 AA.

AC Q9SRJ9; (1)

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE L24373P.

GN EBI OR CG4063.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkeley;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR ENBL; AY061326; AAL28874.1; -

DR FlyBase; FBgn0023444; ebi.

DR GO; GO:0000074; P:regulation of cell cycle; IMP.

DR InterPro; IPR006594; Lish.

DR InterPro; IPR001005; Myb_DNA_binding.

DR Pfam; PF00400; WD40; 3.

DR PRINTS; PR00320; GPROTEINBRPT.

DR ProDom; PD000018; WD40; 1.

DR SMART; SM00667; Lish; 1.

DR PROSITE; PS00678; WD_REPEATS_1; 2.

DR PROSITE; PS00682; WD_REPEATS_2; 3.

DR PROSITE; PS00294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.

SQ SEQUENCE 524 AA; 52529 MW; 6C3D86110BA18D65 CRC64;

Query Match 33.9%; Score 923.5; DB 5; Length 524;

Best Local Similarity 40.5%; Pred. No. 5.8e-56;

Matches 218; Conservative 39; Mismatches 64; Indels 217; Gaps 12;

QY 1 MSISDEVNLFVRYLQESGFHSAPTFGIKSHISQSNINGALVPPAALISITIKGLQYV 60

Db 1 MSFSSDEVNLFVRYLQESGFHSAPTFGIKSHISQSNINGALVPPAALITIKGLLYT 60

QY 61 EAEVINEGTLFDGRPIESLSLIDAMPDV-----VQTRQ----- 96

Db 61 EVENSVDGEV--ARPIEGLSLIDAMPVVKPLPIKTEPGKPGAVDSSAPAGNQN 118

QY 97 -----QAYRDKLA----- 130

Db 119 NAKPEIKIEPGTVAGSAGGNKIAGSTGTSTFDQSASEVDSGNANAGTYAGNG 178

QY 131 ANGE----- 135

Db 179 AGNQASTGGSNSTSTPAGGLAAPGASQKQNSNEAGSSSGNAGNANATSTDDAAS 238

QY 136 ---NG-----AHTIANNHTM----- 148

Db 239 TSTNGNSTSSVBEQPTSGLTAGTVSTSNPDAAASGGASTATGSKAPSGAVTIRVGAQ 298

QY 149 -----MEVDGDVEIIPNKAVLVRGHESEV 172

Db 299 GNNVQSGSSNAQSSAPSGTISSTSGCAGTPAALVPMVDIDENTIPEISKARVLRGHESEV 358

QY 173 FICAWNPFVSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDYPNSKDVTSLD 232

Db 359 FICAWNPFSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDYPNSKDVTSLD 417

QY 233 WNSEGTLLATGSDYGFARIWTKDGNLSTLCQHKGPFIKWKNGKGFILSAGVDKTTII 292

Db 418 WNCDSLLATGSDYGFARIWTKDGNLSTLCQHKGPFIKWKNGKGFILSAGVDKTTII 477

QY 293 WDAHTGKAKQFPFHSAPALVDVQSNNTFASC-----STDMCITV-CK 335

Db 478 WDASTG-----PMHPAICL-----SQCSLIGCGGLADKPGCLLQYGSADTCTCPVPGCK 524

RESULT 13

Q8X1P4

ID Q8X1P4

AC Q8X1P4; PRELIMINARY; PRT; 1356 AA.

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Beta transducin-like protein HET-E2C*4.

GN HET-E.

OS Podospora anserina.

Tue Aug 10 08:05:04 2004

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C*40.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF323584; AAL37300.1; -
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PSS0837; NACHT; 1.
DR PROSITE; PSS0678; WD_REPEATS_1; 10.
DR PROSITE; PSS0082; WD_REPEATS_2; 10.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149694 MW; 572A0B034DCB0D21 CRC64;

Query Match 17.3%; Score 472; DB 3; Length 1356;
Best Local Similarity 31.7%; Pred. No. 5.5e-24;
Matches 111; Conservative 59; Mismatches 154; Indels 26; Gaps 7;

QY 165 LRGHSESVFICAMNPVSDLLASGSGDSTARIMNLSENSTSGTQLVLRHCIREGGQDVPS 224
DB 921 LEHGSSVLSVAFSPDQQRVASGSGDKTIKW---DTASGTGTLEGH---GG----- 968

QY 225 NKDVTSLDWNSGTLATGSGYDGFARIW-TKQGNLASTLGQHKGPALPKWKKGNFILS 283
DB 969 --SWMSVAFSPDQQRVASGSGDKTIKWDTASGTCTQTLLEGHGNVMSVAFSPDQQRVAS 1026

QY 284 AGVDKTTIIWDATGENAKQFPFHSAPALDWDQO-SNNTFASCSSTDMCHVCKLGQDRPI 342
DB 1027 GSDKTKIKIWDATSGTCTQTLLEGHGWVSVAFSPDQQRVASGSDGTIKIWDASGTCT 1086

QY 343 KTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNVHDLQHNKEIYTIKWSPT 402
DB 1087 QTLLEGHDWQSVAFSPDQQRVASGSGDDHTIKIWDASGTCTQTLLEGHGDVMSVAFSPD 1146

QY 403 GFGTNNFNANMLASAFSDFTVRLWDVDRGICHTLTKEQEPYVSVAFSPDGRYLASGSF 462
DB 1147 GQ-----RVASGSDGTIKIWDASGTCTQTLLEGHGWVSVAFSPDQQRVASGSI 1197

QY 463 DKCVHIWNTQTGALVHSYRTGG-IFEVCNAAAGDKVGASADSGVCVLD 511
DB 1198 DGIKIKIWDASGTCTQTLLEGHGWVSVAFSPDQQRVASGSDNTIKIWD 1247

Search completed: August 9, 2004, 16:47:57
Job time : 37.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:39:47 ; Search time 48 Seconds
(without alignments)
3025.613 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLVWIPLVALLIQLVVK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04: *
1: Geneseq1980s: *
2: Geneseq1990s: *
3: Geneseq2000s: *
4: Geneseq2001s: *
5: Geneseq2002s: *
6: Geneseq2003as: *
7: Geneseq2003bs: *
8: Geneseq2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2488	91.5	514	4	AAB95225 Human pro
2	2485	91.4	514	6	ABO07190 Human p53
3	2229	82.0	577	7	ADD14051 Human arc
4	2132	78.4	542	5	ABP41760 Human ova
5	2047	75.3	395	5	ABP51424 Human MDD
6	1721.5	63.3	584	4	ABG211351 Novel hum
7	1715.5	63.1	700	4	ABG60376 Drosophil
8	1137	41.8	208	4	AAO04385 Human pol
9	819	30.1	167	4	ABG21350 Novel hum
10	469	17.3	535	6	ABR52980 Protein s
11	374.5	13.8	414	6	ADA13321 Human int
12	370.5	13.6	411	4	ABG62260 Drosophil
13	367.5	13.5	485	4	AAB68284 Amino aci
14	366.5	13.5	485	4	AAB68516 Human GTP
15	366.5	13.5	485	4	AAB92844 Human pro
16	366.5	13.5	485	5	ABR97306 Novel hum
17	359	13.2	484	4	AAB68282 Amino aci
18	356	13.1	352	3	AAG14893 Arabidops
19	356	13.1	352	3	AAG48119 Arabidops
20	356	13.1	352	7	ADG95040 A. thalia
21	353	13.0	349	3	AAG48120 Arabidops
22	353	13.0	349	3	AAG14894 Arabidops
23	351.5	12.9	409	7	ADSS4235 Human pro
24	350	12.9	339	2	AAW55957 Human cia
25	350	12.9	339	2	AAW58888 Human lip

ALIGNMENTS

RESULT 1

AAB95225
ID AAB95225 standard; protein; 514 AA.
XX AC AAB95225;
XX AC AAB95225;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:17352.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

Ade59693 Human Pro
Ade59705 Human Pro
Ade59697 Human Pro
Ade59709 Human Pro
Ade83378 Human Pro
Ade59701 Human Pro
Ade59713 Human Pro
Ade57861 Human Pro
Ade70002 OPDE 45 k
Ade58488 Rat Prote
Ade83492 Rat Prote
Ade54233 Rat Prote
Ade70005 OPDE 45 k
Ade59486 Drosophil
Aay79678 Drosophil
Aaw06085 Human TAT
Aaw25018 TATA-bind
Aar85868 WD-40 dom
Abb99407 Amino aci
Ada21145 Human sec

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX Sequence 514 AA;

Query Match 91.5%; Score 2488; DB 4; Length 514;
 Best Local Similarity 99.4%; Pred. No. 2.le-224;
 Matches 469; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISSEVNFVLYRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSEVNFVLYRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMEVDGVEIPNKAVLGRGHESEVFICAWNVP 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMEVDGVEIPNKAVLGRGHESEVFICAWNVP 180
 QY 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKNVTSLDWNSGTL 240
 DB 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKNVTSLDWNSGTL 240
 QY 241 ATGSDGFAIWTGDKNLASTLQHKGPFLAKWKKGNFLLSAGVDKTTIWDATGGA 300
 DB 241 ATGSDGFAIWTGDKNLASTLQHKGPFLAKWKKGNFLLSAGVDKTTIWDATGGA 300
 QY 301 KQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIMSKQDNCVHDLQHNKEIYTIKWSPTGPTGNFNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIMSKQDNCVHDLQHNKEIYTIKWSPTGPTGNFNANMLASAF 420
 QY 421 DSTVRLMDVDRGCIHTLTKHQPVSVAFPSPDGRYLASGSPKCVHIWNTQ 472
 DB 421 DSTVRLMDVDRGCIHTLTKHQPVSVAFPSPDGRYLASGSPKCVHIWNTQ 472

RESULT 2
 ABO07190
 ID ABO07190 standard; protein; 514 AA.

XX ABO07190;
 XX ABO07190;
 DT 13-AUG-2003 (first entry)
 XX Human p53 modifying protein, SEQ ID 150.

XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.
 XX W0200299122-A1.

PD 12-DEC-2002.
 XX 03-JUN-2002; 2002WO-US017382.
 XX 05-JUN-2001; 2001US-0296076P.
 PR 10-OCT-2001; 2001US-0328605P.
 PR 15-FEB-2002; 2002US-0357253P.
 XX (EXEL-) EXELIXIS INC.
 XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 XX WPI; 2003-156859/15.
 DR N-PSDB; ACD13365.
 XX Identifying modulators of the p53 pathway for use in treating apoptotic
 PT or cell proliferation disorders, comprises screening for agents that
 PT modulate activity of a human ortholog of genes that modify the p53
 PT pathway in *Drosophila*.
 XX Example 2; Page 458-459; 678pp; English.

CC The invention relates to identifying (M1) a candidate p53 pathway
 CC modulating agent, by contacting an assay system comprising a purified HM
 CC polypeptide (human orthologue of genes that modify the p53 pathway in
 CC *Drosophila*) or nucleic acid with a test agent under conditions, where but
 CC for the presence of the test agent, the system provides a reference
 CC activity, and detecting a test agent-biased activity of the assay system.
 CC Also included are modulating (M2) a p53 pathway of a cell (comprising
 CC contacting a cell defective in p53 function with a candidate modulator
 CC that specifically binds to an HM polypeptide comprising an HM amino acid
 CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
 CC in a mammalian cell (comprising contacting the cell with an agent that
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
 CC a disease in a patient (comprising: (a) obtaining a biological sample
 CC from the patient; (b) contacting the sample with a probe for HM
 CC expression; (c) comparing the results with a control; and (d) determining
 CC whether the comparison indicates a likelihood disease). (M1) is useful
 CC for identifying modulators of the p53 pathway. A probe for HM expression
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
 CC in a patient, where the cancer has greater than 25 % expression level.
 CC Modulators identified by (M1) are useful in a variety of diagnostic and
 CC therapeutic applications, where disease or disorder prognosis is related
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
 CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
 CC the p53 function of the cell, so that the cell undergoes normal
 CC proliferation or progression through the cell cycle. (M2) and (M3) are
 CC also useful for treating defects in the p53 pathway such as angiogenic,
 CC apoptotic or cell proliferation disorders. The present sequence
 CC represents a human p53 pathway modifying protein

XX Sequence 514 AA;

Query Match 91.4%; Score 2485; DB 6; Length 514;
 Best Local Similarity 99.4%; Pred. No. 4e-224;
 Matches 469; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSISSEVNFVLYRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSEVNFVLYRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGENGAAHTIANNHTDMEVDGVEIPNKAVLGRGHESEVFICAWNVP 180
 DB 121 QGSAKNGENTANGENGAAHTIANNHTDMEVDGVEIPNKAVLGRGHESEVFICAWNVP 180
 QY 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKNVTSLDWNSGTL 240
 DB 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKNVTSLDWNSGTL 240

QY 241 ATGSDYGFARIWTKDGNLSTLQHKGPFPALFKWKKGNFILSAGVDKTTIWDHTGSA 300
 DB 241 ATGSDYGFARIWTKDGNLSTLQHKGPFPALFKWKKGNFILSAGVDKTTIWDHTGSA 300
 QY 301 KQGFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQGFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPGTNNNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPGTNNNANMLASASF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVYLAGSPDKCVHIWNTQ 472
 DB 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVYLAGSPDKCVHIWNTQ 472

RESULT 3

ID ADD14051 standard; protein; 577 AA.

XX AC ADD14051;

XX DT 01-JAN-2004 (first entry)

XX DE Human src biomarker polypeptide SEQ ID NO:240.

XX KW predictor set; protein tyrosine kinase activity modulator;
 XX KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 XX KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX OS Homo sapiens.

XX PN WO2003062395-A2.

XX PD 31-JUL-2003.

XX PF 17-JAN-2003; 2003WO-US0001981.

XX PR 18-JAN-2002; 2002US-0350061P.

XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX PI Huang F, Fairchild CR, Lee FY, Shaw P;

XX DR WPI: 2003-636735/60.

XX DR N-PSDS; ADD14646.

XX PT New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX PS Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels
 CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the

CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

Query Match 82.0%; Score 2229; DB 7; Length 577;

Best Local Similarity 85.4%; Pred. No. 5.5e-200;

Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIKGLQYV 60
 DB 52 MSITSDEVNFLVRYLOESGFSHSAFTFGIESHSIQSNINGTLVPPAALISIIKGLQYV 111
 QY 61 BAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAA--- 117
 DB 112 BAEISINEDGTVDGRPIESLSLIDAVMPDVVQTRQAFREKLAQQQASAAAAAATAA 171
 QY 118 -----ASQGSAGKNGENTANGENGAAHTTANNHTDMWVDGVEIIPNKAVILRG 167
 DB 172 ATAATTSGAGVSHQNPKNREATVNGEENRAHSV-NNAKPEIDGVEIIPSKATVILRG 230
 QY 168 HSEVFICAMNPVSDLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGQDVPSNKD 227
 DB 231 HSEVFICAMNPVSDLLASGSDSTARINWLSNSTSGTQLVLRHCIREGGHVDVPSNKD 290
 QY 228 VTSLDWNSGTLTATGSYDGFARITWTKDGNLSTLQHKGPFPALFKWKKGNFILSAGVD 287
 DB 291 VTSLDWNTNGTLTATGSYDGFARITWTEGDLASTLQHKGPFPALFKWKKGNFILSAGVD 350
 QY 288 KTTIWDHTGSAKQGFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 347
 DB 351 KTTIWDHTGSAKQGFPHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQ 410
 QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTG 407
 DB 411 HTNEVNAIKWDPTGMLLASCDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTG 470
 QY 408 NPNANMLASASFDSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRVYLAGSPDKCVH 467
 DB 471 NPNANMLASASFDSTVRLWDIERGVCTHTLTKHQPVPVSVAFSPDGRVYLAGSPDKCVH 530
 QY 468 IWNTQ 472
 DB 531 IWNTQ 535

RESULT 4

ABP41760

ID ABP41760 standard; protein; 542 AA.

XX AC ABP41760;

XX DT 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HNOKM38, SEQ ID NO:2892.

XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.

XX OS Homo sapiens.
 XX PN WO200200677-A1.
 XX PD 03-JAN-2002.
 XX PF 07-JUN-2001; 2001WO-US018569.
 XX PR 07-JUN-2000; 2000US-0209467P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Birse CE, Rosen CA;
 XX DR WPI; 2002-147878/19.
 XX DR N-PSDB; ABQ54837.
 XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 XX PS Claim 11; SEQ ID NO 2892; 2922pp; English.
 XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 542 AA;
 Query Match 78.4%; Score 2132; DB 5; Length 542;
 Best Local Similarity 84.8%; Pred. No. 6.5e-191;
 Matches 395; Conservative 31; Mismatches 26; Indels 14; Gaps 2;
 QY 20 GFHSATFGIKSHISOSNGLALVPPAALISIIQKGLQVVEAEVINEEDGTLFDGRPIE 79
 DB 36 GFHSATFFXIESHISOSNGLTLVPPAALISIIQKGLQVVEAEISINEEDGTVDFGRPIE 95
 QY 80 SLSLIDAVMPDVQTRQAYRDLKACQAAAAA-----ASQGSAXN 126
 DB 96 SLSLIDAVMPDVQTRQAYRDLKACQAAAAAATAATAATTSAGVSHQNPKN 155
 QY 127 GENTANGENGATIANHTDMMEVDGVEIPNKAIVLREHSEVEFCANPNVSDLLAS 186
 DB 156 REATVNGEENRAHSV-NNHAKPMEIDGVEIPSKATVLRGHESEVFCANPNVSDLLAS 214

QY 187 GSGDSTARINWLSNSTSGSTQLVLRHCIREGGDVPNSKDVTSLDWNSGTTLATGSYD 246
 DB 215 GSGDSTARINWLSNSTSGSTQLVLRHCIREGGDVPNSKDVTSLDWNTNGTLLATGSYD 274
 QY 247 GFARITWKDGNLSTLQGHKGPFPALKNKKGNFIISAGVDKTTIWDHAHTGAKQOFPF 306
 DB 275 GFARITWEDGNLSTLQGHKGPFPALKNKKGNFIISAGVDKTTIWDHAHTGAKQOFPF 334
 QY 307 HSAPALVDWQSNNTFASCTDMCIHVCKLQDRPIKTFQGHTEVNAIKWDPTGNLLAS 366
 DB 335 HSAPALVDWQSNNTFASCTDMCIHVCKLQDRPIKTFQGHTEVNAIKWDPTGNLLAS 394
 QY 367 CSDDMTLKINSMKQDNCVHDLQCHNKIYTIKWSPTGPTNPNANMLASASFDSTVRL 426
 DB 395 CSDDMTLKINSMKQEVCIHDLQAHNKIYTIKWSPTGPTNPNANMLASASFDSTVRL 454
 QY 427 WDVDRGICITLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
 DB 455 WDIERGVCITLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 500
 RESULT 5
 ID ABP51424 standard; protein; 395 AA.
 AC ABP51424;
 XX AC
 DT 03-SEP-2002 (first entry)
 XX DT
 DE Human MDDT SEQ ID NO 446.
 XX DE
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiaesthetic; antithrombotic; antitumor;
 KW neuroprotective; antirheumatic; antidiabetic.
 XX OS Homo sapiens.
 XX WO200240715-A2.
 XX 23-MAY-2002.
 XX 06-SEP-2001; 2001WO-US027628.
 XX 05-SEP-2000; 2000US-0229747P.
 XX 05-SEP-2000; 2000US-0229748P.
 XX 05-SEP-2000; 2000US-0229749P.
 XX 05-SEP-2000; 2000US-0229750P.
 XX 05-SEP-2000; 2000US-0229751P.
 XX 05-SEP-2000; 2000US-0230583P.
 XX 06-SEP-2000; 2000US-0230584P.
 XX 06-SEP-2000; 2000US-0230514P.
 XX 06-SEP-2000; 2000US-0230515P.
 XX 06-SEP-2000; 2000US-0230517P.
 XX 06-SEP-2000; 2000US-0230518P.
 XX 06-SEP-2000; 2000US-0230519P.
 XX 06-SEP-2000; 2000US-0230595P.
 XX 06-SEP-2000; 2000US-0230597P.
 XX 06-SEP-2000; 2000US-0230598P.
 XX 06-SEP-2000; 2000US-0230599P.
 XX 06-SEP-2000; 2000US-0230610P.
 XX 06-SEP-2000; 2000US-0230685P.
 XX 06-SEP-2000; 2000US-0230988P.
 XX 06-SEP-2000; 2000US-0230989P.
 XX 07-SEP-2000; 2000US-0230951P.
 XX 07-SEP-2000; 2000US-0231163P.
 XX 07-SEP-2000; 2000US-0231167P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX XX

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PB;
PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
PI Roseberry AM, Gerstin EH, Petalta CH, David NH, Panzer SR, Flores V;
PI Daffo A, Marwaha K, Chen AJ, Chang SC, Au AP, Inman RR;
XX WPI: 2002-527544/56.
DR N-PSDB; ABQ72641.
XX
XX Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
PT AIDS.
XX
XX
PS Claim 14; Page 578; 618pp; English.
XX
XX The invention relates to an isolated human disease detection and
CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound that
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDDT in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDDT in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDDT. (I)
CC or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDDT, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences
XX
SQ Sequence 395 AA;

Query Match 75.3%; Score 2047; DB 5; Length 395;
Best_Local_Similarity 99.5%; Pred.No.3.8e-183;
Matches 389; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISSDEVNVLVRYLQESGFHSFTFGIKHSQSNINGALVPPAALISIIQGLQV 60
DB 5 MSISSDEVNVLVRYLQESGFHSFTFGIESHSQSNINGALVPPAALISIIQGLQV 64

QY 61 EAEVSNEDGTLFDGPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
DB 65 EAEVSNEDGTLFDGPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 124

QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGDIIPPKNKAVLRGHESEVFIQWNPV 180
DB 125 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGDIIPPKNKAVLRGHESEVFIQWNPV 184

QY 181 SLLASGSGDSTARINWLNSENSTSGTLVLRHCRREGQDVPSKNDVTSLDWNSGTL 240
DB 185 SLLASGSGDSTARINWLNSENSTSGTLVLRHCRREGQDVPSKNDVTSLDWNSGTL 244

QY 241 ATGSYDGFARIWTKDGNLSTLGOHKGPFLKWKNGNFIISAGVDKTTIWDHTGEA 300
DB 245 ATGSYDGFARIWTKDGNLSTLGOHKGPFLKWKNGNFIISAGVDKTTIWDHTGEA 304

QY 301 KQOPFHSAPALDWDWOSNNTFASCSDMCIHVCKLGQDREIKTFQGHTEVNAIKWDP 360
DB 305 KQOPFHSAPALDWDWOSNNTFASCSDMCIHVCKLGQDREIKTFQGHTEVNAIKWDP 364
QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLQOHN 391
DB 365 GNLLASCSDDMTLKIWSKQDNCVHDLQAHN 395

RESULT 6
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX
XX AC ABG21351;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #21342.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US008631.
XX
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS85538.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX PS Claim 20; SEQ ID NO 51710; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: the sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 584 AA;

Query Match 63.3%; Score 1721.5; DB 4; Length 584;

Best Local Similarity 63.5%; Pred. No. 2.7e-152;
Matches 338; Conservative 30; Mismatches 35; Indels 129; Gaps 4;
QY 16 LQSGFHSFTGKSHISQSNINGALVPPAALISIIQGLQYVEARVINEEDGTLFPG 75
DB 62 LEVKGFSHSAFTGIESHISQSNINGTLVPPSALISILQGLQYVEARISINKDGTVPDS 121
QY 76 RPTESLIDAVMPDVQFQYRQKLAQQAQAAAAA-----AAASQOGSA 124
DB 122 RPTESLIVAVIPDVVQFQYRQKLAQQAQAAAAA-----AAASQOGSA 181
QY 125 KNGENTANGENGGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCICAWNPVSDLL 184
DB 182 KNGENTANGENGGAHEI--NNHSPKMEIDGDVEIPPNKATVLRGHESEVFCICAWNPVSDLL 240
QY 185 ASGSGDSTARINWLSNSGTSQVLRHCIREGGQDVPSNKDVTSLDWSGTLATGS 244
DB 241 ASGSGDSTARINWLSNSGTSQVLRHCIREGGHVDPSNKDVTSLDWSGTLATGS 300
QY 245 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTIWDATGKAKQOF 304
DB 301 YDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTIWDATGKAKQOF 310
QY 305 PPHSAPALDVDSNNNTFASCSTDMCHVKLGQDRPIKTFQHT----- 349
DB 311 --NAPALDVDSNNNTFASCSTDMCHVKLGQDRPIKTFQHT----- 367
QY 350 -----NEVNAIKWDPT 360
DB 368 KLTMTGKELRPKTPCSDGALLPAGRRPHLLTGDFIPFKRLSALFQNEVNAIKWDPS 427
QY 361 GNLLASCDSDMTLKIMSKMDCNCVHDLQHNKEIYTIKWSPTGPGTNFNANMLASASF 420
DB 428 GNLLASCDSDMTLKIMSKMDCNCVHDLQHNKEIYTIKWSPTGPGTNFNANMLASASF 487
QY 421 DSTVRLWDVDRGCIHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 488 DSTVRLWDVDRGCIHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 539

RESULT 7
ABB60376
ID ABB60376 standard; protein; 700 AA.
XX ABB60376;
XX AC
XX DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 7920.
XX DE
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEXE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL04479.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell

interactions.
XX Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB01840-AB01845) and the encoded proteins (AB057737-AB057742). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 700 AA;
Query Match 63.1%; Score 1715.5; DB 4; Length 700;
Best Local Similarity 52.9%; Pred. No. 1.3e-151;
Matches 349; Conservative 49; Mismatches 71; Indels 191; Gaps 8;
QY 1 MSISDEVNFIYRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISIIQGLQYV 60
DB 1 MSFSDEVNFIYRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISIIQGLQYV 60
QY 61 EARSINEDGTLFDRPIESLIDAVMPDV-----VQTRQ----- 96
DB 61 EVESVSGEDGEV--ARPIEGLSLIDAVMPDVKPLKPIKTEPGKPGAVDSSAPAGNQNN 118
QY 97 -----QAYRDKLA-----QOAAAAAASQSAKQNGENT 130
DB 119 NAKPEIKIEPTGVAGAGGKNIAGSTGTSTPTDQASVDSGNNANAGGTAGNG 178
QY 131 ANGBE----- 135
DB 179 AGNQASTGSGNSTPAGGDLAAPGASQKKSQNSNEAGSSSGNAGNANATSTDDAAS 238
QY 136 --NG-----AHTIANNHTDM----- 148
DB 239 TSTNGNSTSSVEQPTSGLTTPAGGTVSTNPNDAASGGASTATGSKAPSGAVTIRVGAQ 298
QY 149 -----NEVDGDVEIPPNKAVLRGHESEV 172
DB 299 GNVQSGSSNAQSSAPSGTSSSTSGGAGTAPALVPMIDENIEIPSKARVLRGHESEV 358
QY 173 FICAWNPSVILLASGSDSTARINWLSNSGTSQVLRHCIREGGQDVPSNKDVTSLD 232
DB 359 FICAWNPSRDLASGSDSTARINWLSNSDANTN--SNQVLRHCIREGGQDVPSNKDVTSLD 417
QY 233 WNSGTLATGSYDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTII 292
DB 418 WNCDSLLATGSYDGFARITWTKDGNLSTLQGHKGPFPALKWKNKGFIILSAGVDKTTII 477
QY 293 WDAHTGEAKQOFPFHSAPALDVDSNNNTFASCSTDMCHVKLGQDRPIKTFQHTNEV 352
DB 478 WDAHTGEAKQOFPFHSAPALDVDSNNNTFASCSTDMCHVKLGQDRPIKTFQHTNEV 537
QY 353 NAIKWDPTGNLLASCDSDMTLKIMSKMDCNCVHDLQHNKEIYTIKWSPTGPGTNFNAN 412
DB 538 NAIKWCPOGQLLASCDSDMTLKIMSKMDCNCVHDLQHNKEIYTIKWSPTGPGTNFNAN 597
QY 413 LMLASAFDSTVRLWDVDRGCIHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 598 LILASAFDSTVRLWDVDRGCIHTLTKHQBVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 657

RESULT 8
AAO04385
ID AAO04385 standard; protein; 208 AA.
XX AAO04385;
XX AC
XX DT 06-NOV-2001 (first entry)

```

XX DE Human polypeptide SEQ ID NO 18277.
XX DE
XX DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX XX 28-FEB-2000; 2000US-00515126.
XX PR 19-MAY-2000; 2000US-00577409.
XX XX (HYSE-) HYSEQ INC.
XX PA
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AA184316.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX PS Claim 20; SEQ ID NO 18277; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 208 AA;
XX
XX Query Match 41.8%; Score 1137; DB 4; Length 208;
XX Best Local Similarity 99.5%; Pred. No. 4.3e-98;
XX Matches 206; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 213 HCIREGGQVPSNKDVTSLDWNSGTTLLATGSDGFGFARIWTKDGNLASTLGGHKGPIFAL 272
XX Db 1 HCIREGGQVPSNKDVTSLDWNSGTTLLATGSDGFGFARIWTKDGNLASTLGGHKGPIFAL 60
XX
XX QY 273 KNNKGNFLSAGVDKTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSGTDGCIH 332
XX Db 61 KNNKGNFLSAGVDKTTIIWDAHTGEAKQOFPFHSAPALVDWQSNNTFASCSGTDGCIH 120
XX
XX QY 333 VCKLGDRIKTFQGTNEVNAIKDPTGNLLASGDDMTLKIWSMKQDNCVHDJQOENK 392
XX Db 121 VCKLGDRIKTFQGTNEVNAIKDPTGNLLASGDDMTLKIWSMKQDNCVHDJQOENK 180
XX
XX QY 393 EYTIKWSPTGPTGPTNNPNANMLAS 419
XX Db 181 EYTIKWSPTGPTGPTNNPNANMLAS 207
XX
XX RESULT 9
XX ABG21350
XX ID ABG21350 standard; protein; 167 AA.
XX AC ABG21350;

```

```

XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21341.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX XX 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX XX (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AA85537.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 51709; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 167 AA;
XX
XX Query Match 30.1%; Score 819; DB 4; Length 167;
XX Best Local Similarity 85.2%; Pred. No. 2.3e-68;
XX Matches 144; Conservative 13; Mismatches 10; Indels 0; Gaps 0;
XX
XX QY 250 RIWTKDGNLASTLGGHKGPIFALKWNNKGNFLLSAGVDKTTIIWDAHTGEAKQOFPFHS 309
XX Db 1 RIWTKDGNLASTLGGHKGPIFALKWNNKGNFLLSAGVDKTTIIWDAHTGEAKQOFPFHS 60
XX
XX QY 310 PALDQVDMQSNNTFASCSGTDGCIHVCRLGCDHPVKTFQGTNEVNAIKWDPTGNLLASG 369
XX Db 61 PALDQVDMQSNNTFASCSGTDGCIHVCRLGCDHPVKTFQGTNEVNAIKWDPTGNLLASG 120
XX
XX QY 370 DMTLKIWSMKQDNCVHDJQOENKEIYTIKWSPTGPTGPTNNPNANMLA 416
XX Db 121 DMTLKIWSMKQDNCVHDJQOENKEIYTIKWSPTGPTGPTNNPNANMLA 167

```

RESULT 10
 ABR52980
 ID ABR52980 standard; protein; 535 AA.
 XX ABR52980;
 AC ABR52980;
 DT 20-JUN-2003 (first entry)
 XX
 DE Protein sequence #SEQ ID 825.
 DE
 KW Multiprotein complex; eukaryote; drug target; diagnosis.
 XX
 OS Saccharomycetes cerevisiae.
 XX
 PN EP1258494-A1.
 XX
 PD 20-NOV-2002.
 XX
 XX 20-DEC-2001; 2001EP-00130253.
 XX
 XX 15-MAY-2001; 2001EP-00111774.
 XX
 PA (CELL-) CELLZONE AG.
 XX
 PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzloch M, Schultz JD, Superti-Furga GD;
 XX
 DR WPI; 2003-250078/25.
 DR
 DR N-PSDB; ACC61022.
 XX
 XX
 PT New isolated protein complexes useful for diagnosing a disease or
 PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 PT disorder.
 XX
 PS Disclosure; SEQ ID NO 825; 17pp + Sequence Listing; English.
 CC
 CC The invention relates to multiprotein complexes from eukaryotes. Proteins
 CC of the invention and DNA sequences encoding them are given in records
 CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
 CC obtainable by using a protein as a bait and isolating the set of proteins
 CC which is attached thereto from cells. Such protein complexes may comprise
 CC up to 30 distinct proteins. Protein complexes of the invention are useful
 CC for diagnosing a disease or disorder, or as a target for an active agent
 CC of a pharmaceutical, preferably a drug target in the treatment or
 CC prevention of a disease or disorder. Note: The sequence data for this
 CC patent is not represented in the printed specification, but is based on
 CC sequence information supplied by the European Patent Office. The complete
 CC document is available on CD-ROM
 XX
 SQ Sequence 535 AA;
 Query Match 17.3%; Score 469; DB 6; Length 535;
 Best Local Similarity 25.8%; Pred. No. 1.1e-34;
 Matches 140; Conservative 91; Mismatches 230; Indels 82; Gaps 15;
 1 MSITSSEVNFVLYRLQESGFHSA-----FTFGIKSHISQSNINGALVPPAAL 49
 1 MSITSSEVNFVLYRLQESGFHSA-----FTFGIKSHISQSNINGALVPPAAL 50
 50 ISIIQKGLQVRAEVINEDG-----TLFDGRPIESLIDAVMPDVQVTOQQAYRKLAAQ 106
 51 VNLVQRIYTESLWVDSKDISALNEHLSDFNLVQALQID-----KEKFPPI 101
 107 QAAAAAASQQAQSGKNGTANGENGATIANHNDMMVEDVDVEIPNKAVALR 166
 102 SSGRTLETNSNKRAGEDGASTVERETQEDDTNSIDSD--DLDFVKI--LKEIV-- 155
 167 GHESEVFCANVPSP--LLASGSGDSTARINWSENSTSGSTG-----LVLRHCIREGG 219
 156 -KLDNIVSSTWNPDLDESILAYGEKNSVARLARIVETDQEGKYWKLTIIAELRHPPALSA 214

QY 220 QDVPSNKDVTSLDWNSEGTLLATGSDGFARIWTKGNLASTLQGHKGPFLAKWKKGN 279
 DB 215 SSGTKTNTQVTCCLAWSHDGNSIVTGVENGELRLNKTGALLNLNFRAPITVSVMKNDGT 274
 QY 280 FILSAGVDKTTIWDATHTGEAKQOFPF-----HSAP-----ALDVMQSNNTFAS 324
 DB 275 HIIISMDVENVTILNNVISGIVMWHFELKGTGSSINAENHSGSLGVDVWDDXKEVI 334
 QY 325 CSTDMCHVCKLQGDRIPIKTFQHTNEVNAIKWDPTGNLLASCSDDMTLKIMSKQDNCV 384
 DB 335 PGPKGAIFVQITEKTPTKLIGHHGPISVLEFENDTKLLLSASDDGTLRIMHGGNGNSQ 394
 QY 385 HDLQCHNKEIYTIKWSPGPTGNPNANMLASASPDSTVRLWDVDRGICHTLTQHQP 444
 DB 395 NCFYGHSSQSVASWV-----GDDKVICSDMGSVRLWSLKQNTLLAISVDGVP 444
 QY 445 VYSVAFSPDGRYLASGSFDFKCVHIWNTQVCLHYLNGQVLLNLG-RSICLYTLPHHLVVP 503
 DB 445 IFAGRISQDGQKYAVAFPMQGVNVD---LKKLNSKRSRLYGNRDLNPLP-----IP 495
 QY 504 LVA 506
 DB 496 LVA 498
 RESULT 11
 ADA13321
 ID ADA13321 standard; protein; 414 AA.
 XX ADA13321;
 AC ADA13321;
 DT 06-NOV-2003 (first entry)
 XX
 DE Human intracellular signalling molecule INTSIG-3, SEQ ID NO:3.
 XX
 KW Human; intracellular signalling molecule; INTSIG;
 KW cell proliferative disorder; cancer; atherosclerosis;
 KW autoimmune disorder; inflammatory disorder; infection;
 KW neurological disorder; developmental disorder; endocrine disorder;
 KW cystostatic; antiarteriosclerotic; nootropic; neuroprotective;
 KW cerebroprotective; immunosuppressive; antiinflammatory; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO2003031568-A2.
 XX
 PD 17-APR-2003.
 XX
 PF 16-AUG-2002; 2002WO-US026322.
 XX
 XX 17-AUG-2001; 2001US-0313245P.
 PR 24-AUG-2001; 2001US-0314751P.
 PR 31-AUG-2001; 2001US-0316752P.
 PR 31-AUG-2001; 2001US-0316847P.
 PR 14-SEP-2001; 2001US-0322186P.
 PR 28-SEP-2001; 2001US-0326390P.
 PR 12-OCT-2001; 2001US-0328952P.
 PR 19-OCT-2001; 2001US-0345468P.
 PR 12-APR-2002; 2002US-0372499P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Lu DM, Swarnakar A, Tang YT, Griffin JA, Emerling BM;
 PI Forsythe IU, Yao MG, Ramkumar J, Richardson TW, Becha SD, Lee EA;
 PI Warren BA, Lehr-Mason PM, Baughn MR, Li JX, Duggan BM, Gietzen KJ;
 PI Lal PG, Borowsky ML, Ison CH, Thangavelu K, Xu Y, Lee S;
 PI Elliott VS, Spague WW, Azimzai Y, Hefalia AJA, Ding L, Nguyen DB;
 PI Honchell CD, Luo W, Chawla NK, Marquis JP, Jackson JL, Tran UK;
 XX
 DR WPI; 2003-393436/37.
 DR N-PSDB; ADA13366.
 XX
 PT New human intracellular signalling molecules (INTSIG), useful for

PT diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
 PT infections.

XX Claim 1; Page 225-226; 367pp; English.

XX The invention relates to 45 human intracellular signalling molecules
 CC (INTSIG), INTSIG-1 to INTSIG-45, and the cDNAs encoding them (ADAL319-
 CC ADAL3408). The invention also encompasses expression constructs, host
 CC cells and transgenic organisms comprising an INTSIG nucleic acid sequence
 CC ; the recombinant preparation of a INTSIG; an antibody against a INTSIG;
 CC methods of detection of INTSIG proteins or nucleic acids; a micro-array
 CC containing INTSIG nucleic acids; methods of screening compounds for their
 CC ability to modulate INTSIG activity or expression; and pharmaceutical
 CC compositions comprising an INTSIG protein, an INTSIG antibody, an INTSIG
 CC agonist or INTSIG antagonist. The INTSIG proteins, nucleic acids or
 CC compositions comprising them are useful in diagnosing, treating or
 CC preventing a variety of disorders, including cell proliferative disorders
 CC (e.g., cancer or atherosclerosis); autoimmune/inflammatory disorders
 CC (e.g., AIDS, allergies, anaemia, asthma, bronchitis, gout, multiple
 CC sclerosis, osteoarthritis, rheumatoid arthritis, or systemic lupus
 CC erythematosus); neurological disorders (e.g., epilepsy, stroke,
 CC Alzheimer's disease, dementia, or Parkinson's disease); viral, bacterial,
 CC fungal, parasitic, protozoan or helminthic infections; neurological
 CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, dementia, or
 CC Parkinson's disease); developmental disorders (e.g., achondroplastic
 CC dwarfism, Cushing's disease, hypothyroidism or hydrocephalus); endocrine
 CC disorders (e.g., disorders of the hypothalamus or pituitary gland or
 CC diabetes); reproductive disorders; or vesicle-trafficking disorders
 CC (e.g., cystic fibrosis or hypercholesterolaemia). The present sequence
 CC represents an INTSIG protein of the invention.

XX Sequence 414 AA;

Query Match 13.8%; Score 374.5; DB 6; Length 414;

Best Local Similarity 28.3%; Pred. No. 5.7e-26;
 Matches 99; Conservative 49; Mismatches 137; Indels 65; Gaps 10;

QY 129 NTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHSEVFICAW-NEVSDLLASG 187
 DB 120 DTASGEELN-----TLEGRNVVYALAEENPEYGDKIATG 153
 QY 188 SGDSTARINWLNENSTSGTQVLVLRHCIREGGQDVPSKDVTSLDWNSGTLATGSDYG 247
 DB 154 SPDKTKLMSVETG-----CYHTF-----GHTAEIVCLSFNPOSTLVATGSMGT 199
 QY 248 FARITWDCGNLASTLQGHKGPFPALKWKKGNFILLSAGVDKTTIILWDAHTGEAKQFPFH 307
 DB 200 TAKLWDIQNGEELTLRGHSAEIIISFNSTSGRIITGSPDHTVVVWDADTGRKNVILICH 259
 QY 308 SA--PALVDVWOSNNTFASCS---TDMCHVCKL---GQDRPIKTFQGHNEVNAIKWDP 359
 DB 260 CAEISSASFNW-----DCSLILTGSMDTKCKLWDATNGKCVATLTGHDDEILDSCFDY 312
 QY 360 TGNLLASCSDDMTLKIWSMKQDNCVHDLOQHKEIYTIKWSPTGCTNNPNANLMLASAS 419
 DB 313 TGKLTATAGDGTARIFSAATKRAKLEGHGEISKISF-----NPQGNHILLTGSS 364
 QY 420 FDSVRLWVDVRGICHTHTTKHQEPVSVAFSPDGRYLASGSDKCVHIW 469
 DB 365 -DKTARIWDAQTGCQCLQVLEHTDEIFSCAFYKGNVITGSKDNTCRIW 413

RESULT 12

ID ABB62260
 AC ABB62260;
 AC ABB62260;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 13572.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06363.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 13572; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 411 AA;

Query Match 13.6%; Score 370.5; DB 4; Length 411;

Best Local Similarity 22.4%; Pred. No. 1.3e-25;
 Matches 115; Conservative 77; Mismatches 158; Indels 163; Gaps 14;

QY 6 DEVNFLVRYLQESGFGSHSAFTGFIKSHIS---QSNINGALVPPAALISITIKGLQVVEA 62
 DB 11 EELNQAADYLGSGNYADSLSTFEKADLSTEVEKFGGLLEKKWTSVIRLQKKWLEA 70
 QY 63 EVSINEGDTLPDGRPIESLSLIDAVMPDVVQTRQQAYRDKLAQQQAAAAAASQQG 122
 DB 71 KLTEAE-----KEVIE----- 81
 QY 123 SAKGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHSEVFICAWNPVSD 182
 DB 82 ---GATKTKRTGEM-----TPRPEK-FSLTGHASITRVIFHFIFA 121
 QY 183 LLASGSDSTARINWLN-----ENSTSGTQVLVLRHCIREGGQDVPSKDVTSLDWNSG 238
 DB 122 LMSASADATIRIWDFTGEYERSLKHTDSV-----QDVA-----FDAQGX 163
 QY 239 LLATGSDYDGFARIW--TKDGNLASTLQGHKGPFPALKWKKGNFILLSAGVDKTTIILWDAH 296
 DB 164 LLASCSADLSIKLWDFQOSTECIKTMGHCHDNVSSVAFVPAQDYVLVSASDRITKMEVA 223
 QY 297 TGEAKQOFPFHSAFALPDVDMQSNNTFASCSDTMCHIVCKLQDRPIKTFQGHNEVNAIK 356
 DB 224 TG-----YC-----VKTGTGHEWVEMVR 242
 QY 357 WDPTEGNLLASCSDDMTLKIWSMKQDNCVHDLOQHKEIYTIKWSP-----TGPGTNN 408
 DB 243 VHIEGSIATCSNDQTIIRVMTNSKDKVELRDEHTVECIAWAPAAAAAINEAAGADN 302
 QY 409 PNAN---LMLASASFDSVRLWVDVRGICHTHTTKHQEPVSVAFSPDGRYLASGSDK 465

Db 303 KKGHGQPLASGRDKTIRINDVSVGLCLLTLGSHDNVVRGLAFHPGKYLVSASDDKT 362
 QY 466 VHIWNTQVCLHYLNGQVLLNGLRSICLYTLPHH 498
 Db 363 IRW-----DLNKRCKMITYAH 380

RESULT 13
 AAB68284
 XX AAB68284 standard; protein; 485 AA.
 AC AAB68284;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE Amino acid sequence of a human Tzap gene cDNA clone.
 XX
 KW Tzap; T cell activation; immune response; transplant rejection;
 KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosis;
 KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;
 KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;
 KW lepramatosis; gastritis; skin tumour; adrenal tumour; lung tumour;
 KW wound healing; growth disorder; inflammatory disease; infectious disease.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 379
 FT /note= "unspecified residue encoded by NTC"
 XX
 XX W0200132614-A2.
 XX
 PD 10-MAY-2001.
 XX
 XX 30-OCT-2000; 2000WO-EP010670.
 XX
 XX 01-NOV-1999; 99US-0162675P.
 PR 25-FEB-2000; 2000US-0185016P.
 XX
 XX (UTKU/) UTKU N.
 PA
 XX Utku N;
 PI
 XX WPI; 2001-316410/33.
 DR N-PSDB; AAF85075.
 DR
 XX Novel polynucleotide encoding Tzap protein involved in T cell activation,
 PT useful for diagnosing and treating diseases involving T cell activation,
 PT for treating organ transplantation rejection, rheumatoid arthritis.
 XX
 XX Example; Page 56-57; 68pp; English.
 PS
 XX The present sequence represents a Tzap protein. The Tzap gene is involved
 CC in T cell activation, and in the modulation of immune responses. Tzap
 CC polynucleotides and polypeptides are useful for diagnosing or treating
 CC acute and chronic diseases involving T cell activation and Th1 and Th2
 CC immune response, for the treatment of acute and chronic rejection of allo
 CC - and xeno organ transplants and bone marrow transplantation, for the
 CC treatment of rheumatoid arthritis, lupus erythematosis, multiple
 CC sclerosis, encephalitis, vasculitis, diabetes mellitus,
 CC pancreatitis, gastritis, thyroiditis, for the treatment of malignant
 CC disorders of T, B or NK cells, for the treatment of asthma, lepramatosis,
 CC Helicobacter pylori associated gastritis or for the treatment of skin
 CC tumours, adrenal tumours or lung tumours, wound healing, growth disorders
 CC inflammatory and/or infectious diseases
 XX
 XX Sequence 485 AA;

Query Match 13.5%; Score 367.5; DB 4; Length 485;
 Best Local Similarity 26.8%; Pred. No. 3.4e-25;
 Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAMNPYDILLASGSDSTARIWNLSEN-----STSGSTQVLVLRHCIREGGQ 220
 Db 110 LEHSEAVISVAFSPGKYLASGSDTTVRFWDLSTETPHFTCKGHRHWVLSISWSPDGK 169
 QY 221 DVPS-----NKDYTSLDW-----NSEGTLATGSYDGFAR 250
 Db 170 KLASGCKNGOILLWDPSTGKQVGRITLAGHKSMTGLSWEPLHANPECRYVASSSKDGSVR 229
 QY 251 IW-TKQGNLASTLQOHKGPFPALKWKKGNFILLSAGVDKTIITWDAHTGEAKQPFPH-- 307
 Db 230 IMDTTAGRCRILITGHTQSVTCLRWGGDG-LLYSASQDRTIKVWRAHDGVLCTLQGHG 288
 QY 308 --SAPALDVM-----QSNNTFASCS 326
 Db 289 WNTMALSTDYALRTGAFEPABASVNPQDLQSGIQELKERALSRYNLVRGQGERLVSGS 348
 QY 327 TDMCIHVYKLGQD-RPIKTFQGHTEVNAIKWPTGNLASCDDMTLKWSKKQNCVH 385
 Db 349 DDFTLFLWSPAEDKKPLTRMTGHQALINQVXFSFDSRIVASASFDKSIKLMWDGRTGKYLA 408
 QY 386 DLQOHKXEIYTIKWSPTGPTNNPNANMLASASFSSTVRLVDVDRGICIHITLTKEQEPV 445
 Db 409 SLRGHVAAVQIAWSAD-----SRLVSGSSDSTLKVDVKAQKLANDLPGHADDEV 459
 QY 446 YSAFSPDGRYLASGSPDKCVHIW 469
 Db 460 YAVDWSPDGQVAVSGGKDKCLRIW 483

RESULT 14
 AAB68516
 ID AAB68516 standard; protein; 485 AA.
 XX
 AC AAB68516;
 XX
 DT 19-APR-2001 (first entry)
 XX
 DE Human GTP-binding associated protein #16.
 XX
 KW Human; guanosine triphosphate binding associated protein; GTP; GRAP;
 KW inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
 KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
 KW osteoporosis; psoriasis.
 XX
 OS Homo sapiens.
 XX
 XX W0200105970-A2.
 PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-US019698.
 XX
 PR 19-JUL-1999; 99US-0144595P.
 PR 23-AUG-1999; 99US-0150460P.
 PR 15-OCT-1999; 99US-0159849P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 PI Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
 PI Reddy R, Yang J, Baughn MR, Lu DAM, Azimzai Y, Patterson C;
 XX
 XX WPI; 2001-091972/10.
 DR N-PSDB; AAF58316.
 XX
 XX New guanosine triphosphate-binding associated proteins (GBAP) and their
 PT encoding nucleic acids, useful for treating and/or diagnosing diseases
 PT associated with GBAP expression, such as cancer, diabetes and asthma.
 XX
 XX Claim 1; Page 133-134; 233pp; English.
 XX
 XX The present invention relates to novel human guanosine triphosphate (GTP)
 CC -binding associated proteins (GBAPs; AAB68501-AAB68566) and their coding
 CC sequences (AAF58301-AAF58366). The proteins and coding sequences of the

CC present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis

XX Sequence 485 AA;

Query Match 13.5%; Score 366.5; DB 4; Length 485;
Best Local Similarity 26.8%; Pred. No. 4.2e-25;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAWNPVSDLLASGSGDSTARINWLSN-----STSGTQLVLRHCIREGQ 220
DB 110 LEHSEAVISVAFSPGKYLASGSDTIVRFDLSTPHFTCKGHRHWLWSISWSPDGK 169
QY 221 DVPS-----NKDVTSLDW-----NBSGTLATGSDYDGFAR 250
DB 170 KLASGCKNGQILLWDPSGKQVGRTRLAGHSHKWTGLSWEPLHNPCECRYVASSSKDGSVR 229
QY 251 IW-TKDGNIASLTGQHKGPFAKWNKGNFILSAGVDKTTIWDHTGEAKQQPFPH-- 307
DB 230 IWDTTAGRCERILTGHTQSVTCLRWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288
QY 308 --SAPALDWDW-----NKDVTSLDW-----QSNNTFASCS 326
DB 289 WNTMALSTDYALRTGAFEPASVNPQDLQGSQELKERALSRYNLVRGQPERLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVH 385
DB 349 DDFTFLWSPAEDEKPLTRMTGHOALINQVLFSPDSRIVASASFDKSIKLDWGRGKYLA 408
QY 386 DLOOHNKKEITYIKWSTGPTGNTNPNANMLASASFDSTVRLWDVDRGICHTLTKHOEPV 445
DB 409 SLRGHVAAYQIAWSAD-----SRLVSGSSDSTLKWDVKAQKLANDLPGHAD 459
QY 446 YSVAFSPDGRYLASGSDKCVHIW 469
DB 460 YAVDWSFDGQRVASGKDKCLRIW 483

RESULT 15

AAB92844

ID AAB92844 standard; protein; 485 AA.

XX AAB92844;

XX 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11400.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 11400; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification, where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the one of the 5602 oligonucleotide sequences comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 485 AA;

Query Match 13.5%; Score 366.5; DB 4; Length 485;

Best Local Similarity 26.8%; Pred. No. 4.2e-25;

Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;

QY 165 LRGHSEVFICAWNPVSDLLASGSGDSTARINWLSN-----STSGTQLVLRHCIREGQ 220

DB 110 LEHSEAVISVAFSPGKYLASGSDTIVRFDLSTPHFTCKGHRHWLWSISWSPDGK 169

QY 221 DVPS-----NKDVTSLDW-----NBSGTLATGSDYDGFAR 250

DB 170 KLASGCKNGQILLWDPSGKQVGRTRLAGHSHKWTGLSWEPLHNPCECRYVASSSKDGSVR 229

QY 251 IW-TKDGNIASLTGQHKGPFAKWNKGNFILSAGVDKTTIWDHTGEAKQQPFPH-- 307

DB 230 IWDTTAGRCERILTGHTQSVTCLRWGGDG-LLYSASQDRTIKVRAHDGVLCTLQGHGH 288

QY 308 --SAPALDWDW-----NKDVTSLDW-----QSNNTFASCS 326

DB 289 WNTMALSTDYALRTGAFEPASVNPQDLQGSQELKERALSRYNLVRGQPERLVSGS 348

QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVH 385

DB 349 DDFTFLWSPAEDEKPLTRMTGHOALINQVLFSPDSRIVASASFDKSIKLDWGRGKYLA 408

QY 386 DLOOHNKKEITYIKWSTGPTGNTNPNANMLASASFDSTVRLWDVDRGICHTLTKHOEPV 445

DB 409 SLRGHVAAYQIAWSAD-----SRLVSGSSDSTLKWDVKAQKLANDLPGHAD 459

QY 446 YSVAFSPDGRYLASGSDKCVHIW 469

DB 460 YAVDWSFDGQRVASGKDKCLRIW 483

Search completed: August 9, 2004, 16:45:26

Job time : 50 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 9, 2004, 16:41:43 ; Search time 13.333 Seconds
(without alignments)
3708.183 Million cell updates/sec
Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHELWVPLVALIELLVK 514
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	17.5	1526	2 AC2239	WD-40 repeat prote
2	472	17.4	535	2 S48268	probable membrane
3	458.5	16.9	1258	2 A12155	WD-repeat protei
4	451.5	16.6	1683	2 AF2071	WD-40 repeat prote
5	444.5	16.4	1227	2 AE1810	WD-40 repeat prote
6	429.5	15.8	1189	2 A12493	WD-repeat protei
7	426	15.7	1708	2 AE1866	WD-40 repeat prote
8	420.5	15.5	1356	2 T18521	beta transducin-li
9	408	15.0	1711	2 A18442	WD-40 repeat prote
10	404.5	14.9	934	2 AG1889	WD-40 repeat prote
11	395	14.5	1747	2 AC1842	WD-40 repeat prote
12	394.5	14.5	1693	2 S76086	WD-40 repeat prote
13	382.5	14.1	564	2 T40883	beta transducin-li
14	381	14.0	1551	2 AE2410	WD repeat protei
15	369	13.6	676	2 AH2195	WD-repeat protei
16	368.5	13.6	304	2 AG1837	hypothetical prote
17	366	13.5	502	2 T41148	WD-40 repeat prote
18	365	13.4	1189	2 AH2154	WD-repeat protei
19	354	13.0	265	2 AF1890	WD-repeat protei
20	349.5	12.9	410	2 S48052	platelet-activatin
21	344.5	12.7	559	2 AD2202	hypothetical prote
22	342	12.6	409	2 S36113	LIS-1 protein - hu
23	341.5	12.6	677	2 AE1861	serine/threonine k
24	339.5	12.5	786	2 AG2375	WD-40 repeat-prote
25	337	12.4	515	2 S19487	hypothetical prote
26	332	12.2	323	2 T02617	hypothetical prote
27	331	12.2	437	2 S05357	hypothetical prote
28	326.5	12.0	787	2 T00798	hypothetical prote
29	324	11.9	342	2 AE2490	WD-repeat protei

WD-repeat protein
probable WD-repeat
PF20 protein, micr
hypothetical prote
probable sulfat me
WD-40 repeat prote
hypothetical prote,
transcription init
TATA box-binding p
WD-40 repeat regul
hypothetical prote
hypothetical prote
MET30 protein - ye
beta transducin-li
hypothetical prote
hypothetical WD-re

ALIGNMENTS

RESULT 1
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA000019; PID:G17132599; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 17.5%; Score 475; DB 2; Length 1526;
Best Local Similarity 31.0%; Pred. No. 7.1e-26;
Matches 119; Conservative 63; Mismatches 138; Indels 64; Gaps 10;
QY 136 NGAHTIANNHDMVEVDGVEIPENKAV-VLRGHESEVFICAWPVSDLLASGSDSTAR 194
Db 1127 NGV-TLNGSSDQIVRLWD--ISSKCLYTLQGHNNWNAVAFSPDGAATLASGGGQTVR 1183
QY 195 IWNLSSENS---TSGSTQLVLRHCIREGGQDVFPSNKDVTSLDNSEGTLLATSGYDGFAR 250
Db 1184 LWDISSSKCLYILQGHTSW-----VNSVFNPDGSLASGSDQTVR 1225
QY 251 IW-TKDNGLASTLQGHKGPIFALKNKKNKGNFILSAGVDKTIIWDAHTGAQKQFPFHS 309
Db 1226 LWEINSSKCLCTFGHTSWNSVFNPDGSMASGSDKTVRLWDISSSKCLHTFQGH 1284
QY 310 PALDVDVQSNNTF-----ASCSTDMCIHVCKLQDRIKTFQGHNEVNAIKWDPTGN 362
Db 1285 -----NNVNSVAFNPDGSMLASGSDQTVRLWISSSKCLHTFQGHISVSVTFSPDGT 1339
QY 363 LLASCSDDMTLKWSMKQDNVHDLQHNKEIYTIKWSPTGPTGNTNPNANMLIASAFDS 422
Db 1340 MLASGSDQTVRLWISSSGECLYTLFGHTNWGVSIFSPDG-----AILASGSDQ 1390
QY 423 TVRLWDVDRGICHTLTKEHPVYVAFSPDGRYLAGSPDKCVHVIWNTQVCLHYLNGQV 482
Db 1391 TVRLWISSSKCLYTLQGHNNWVGVSIFSPDG-----AILASGSDQTVRLW 1437
QY 483 LLNLGRSICLYTLPHHLVVIPLVA 506
Db 1438 --NISSGECLYTLHGHSVRSVA 1459

```
Db 445 IFAGRISQGGKYAVAFMDGQVNVYD-----LKLNSKSRSLYGNRGDILNPLP-----IP 495
Qy 504 LVA 506
Db 496 LYA 498

RESULT 3
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
A:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2071
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1258 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE74499.1; PID:gl7131893; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2800

Query Match 15.9%; Score 458.5; DB 2; Length 1258;
Best Local Similarity 32.0%; Pred. No. 8.5e-25;
Matches 105; Conservative 43; Mismatches 101; Indels 79; Gaps 8;

Qy 165 LRGESEVFICAWNPVSDLLASGSDSTARIMNLSENSTSGTQLVLRHCIREGGQDVPS 224
Db 974 LEHTDFYIGAFSPDSQTLASASTDSVRLNI---STGQCQILLEH----- 1019

Qy 225 NKDVTSLDW-----NSEGTLATSYDGFARIWT-XDGNLSTLQGHKGPFIALKWKK 277
Db 1074 GQLLASASADQSVRLWD-----CCTGRCVGI----- 1099

Qy 338 QDRPIKTFQGHTEVNAIKWDPCTGNLLASGDDMTLKIWSMKQDNCVHDLQHNKEIYTI 397
Db 1100 -----LRGHSNRYSAIFSPNGEIIATCSTDQTVKIWDWQCGKCLKTLTGHTNWFVDI 1152

Qy 398 KWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTKHQPVPYVAFSPDGRYL 457
Db 1153 AFSPDGK-----LLASASHDQTVRIWDVNTGKCHICIGTHLVSSVAFSPDGEV 1203

Qy 458 ASGSFDCVCHWN--TVCLHLNGQVL 483
Db 1204 ASGSQDQTVRIWNVKTEGCLILAKRL 1231

RESULT 4
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
A:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2071
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2071
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1683 <KUR>
```

```
probable membrane protein YBR103w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR0832
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S48268; S45971; S44683
R:Wannhaupf, G.; Stucka, R.; Ehmlé, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357; PMID:7900426
A:Accession: S48268
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-535 <MAN>
A:Cross-references: EMBL:X78993; NID:G476045; PIDN:CAAS5606.1; PID:G476059
R:Feldmann, H.; Mannhaupf, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45971
A:Molecule type: DNA
A:Residues: 1-535 <FE2>
A:Cross-references: EMBL:Z35972; MIPS:YBR103w
C:Genetics:
A:Gene: SGD:SIF2
A:Cross-references: SGD:S0000307
A:Map position: 2R
A:Superfamily: unassigned WD repeat proteins; WD repeat homology
C:Keywords: transmembrane protein
F:355-388/Domain: WD repeat homology <W1>
F:397-429/Domain: WD repeat homology <W2>
F:434-450/Domain: transmembrane #status predicted <TM>

Query Match 17.4%; Score 472; DB 2; Length 535;
Best Local Similarity 25.8%; Pred. No. 2.8e-26;
Matches 140; Conservative 91; Mismatches 230; Indels 82; Gaps 15;

Qy 1 MSITSDVNFVLYRLOESGFHSA-----FTFGIKSHISQSNINGALVPPAAL 49
Db 1 MSITSEELNYLIWYQCQEMGHEVSALALQDETRLFDEFKYEKHI-----PLGTL 50

Qy 50 ISIIQKGLQVHEARVSNEDG---TLFDGRPIESLSLIDAVMPDVVQTQQAVDQLAQ 106
Db 51 VNLVQRGILYTELMVDSKGDTSALNEHLSDFNLVQALQID-----KEAFPEI 101

Qy 107 QAAAAAASQGSQAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPEKNKAVLR 166
Db 102 SSGRFTLENTSESNKAGEDGASIVERTQEDDTNSIDSSD--DLGDFVKI--LKEIV-- 155

Qy 167 GHSEVFICAWNPVSD--LLASGSDSTARIMNLSENSTSGTQ-----LVLRHCIREGG 219
Db 156 -KLDNIIVSTWNPDESILAYGEKNSVARLARIIVETDQGGKTKWKTITAELRHPFALSA 214

Qy 220 QDVPSNKDVTSLDNNSGTLATSGYDGFARIWTKGNLSTLQGHKGPFIALKWKKGN 279
Db 215 SSGKITNQVTLAWSHGNSIVIGVENGELNKNKTGALLNLVNFHRAPIVSVKWNKOGT 274

Qy 280 FILSAGVDKTTIWDHAHTGAKQFPF-----HSAP-----ALVDWQSNNTFAS 324
Db 275 HIISMDVENVTILNVIISGTVQHFELKETGGSSINAENHSGDGLGVDEWVDDDKFVI 334

Qy 325 CSTDMCHVCKLQGRPIKTFQGHTEVNAIKWDPCTGNLLASGDDMTLKIWSMKQDNCV 384
Db 335 PGPKGALFVYQITETKTGKLIHGHPISVLEFNDTNKLLSASDDGTLRIHGGGNSQ 394

Qy 385 HDLQHNKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTKHQP 444
Db 395 NSFYGHQSIVSAGW-----GDDKVISCMDSGVRLWSLKNTLLALSIVDGV 444

Qy 445 VYSVAFPDGGRYLAGSFDKCVHITWTVQVCLHYLNGQVLNLG-RSICLYTLPHHLVWP 503
Db 445 VYSVAFPDGGRYLAGSFDKCVHITWTVQVCLHYLNGQVLNLG-RSICLYTLPHHLVWP 503
```

A:Cross-references: GB:BA0000019; PIDN:BAB73823.1; PID:g17131215; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2124

Query Match 16.6%; Score 451.5; DB 2; Length 1683;
Best Local Similarity 28.1%; Pred. No. 4.1e-24;
Matches 123; Conservative 75; Mismatches 167; Indels 73; Gaps 11;
QY 64 VSINEDG-TLFDGRPIESLSIDAVMPDVVQTRQAYRDKLAQQOAAAAAASQOQ 122
DB 1245 LSFSPDGKTLASASADKTKLRIADGKLVT-LKGHNSVWVNFSSDGGKAIASASDN 1303
QY 123 SAKNGENTANGENGARTIANNHTDMVEVDGVEIPPKNKAVVLRGHESEVFCANVPVS 182
DB 1304 TIK-----LWNRHGIELE-----TFTCHSGGVAVNPLPDSN 1335

QY 183 LLAASGGSTARIWNLSENSTSGSTQVLVLRHCIREGGQDVPSNKDVSTLDWNSGTTLLAT 242
DB 1336 IIASASLNTIRLMQRPLISPL-----EVLAGNSGVYAVSFILHDSIIAT 1380

QY 243 GSYDGFARIW-TKDGNLSTLQCHKGPFLAKWKKGNFILLSAGVDKTTIILDAHTGEAK 301
DB 1381 AGADGNICLWHSQDSGLKTLPGNKA-IYGISFTPGGLIASANADKTKLWRVDRGKAL 1439

QY 302 QOFPFHSAPALVDWQ-SNNTFASCSTDMCIHVCKLQDRIPIKTPQGHNTNEVNAIKWDPT 360
DB 1440 KTLIGHDNEVKNVNSPDGKTLASASRDNVTVKLVNSDGKPKTKLKGHTDEVFWVSFPD 1499

QY 361 GNLASCSDMTLKTWSMKQDCNVHDLOQHNKEIYTIKWSFTGPGTNNPNANMLASASF 420
DB 1500 GKIIASASADKTIRLWDSFSGNLKSLPAHNDLVVSVNFPDGS-----MLASTSA 1550

QY 421 DSTVRLWDVDRGICHTLTKHQEPVVSFAFSPDGRYLAGSPDKCVHIWNTQVCLHYLNG 480
DB 1551 DKTVKLWNSHDLHLTFSGHNNVYSSFSFDGRYIASASEDKVVKLWQ-----IDG 1603

QY 481 QVLLNLGRSICLYTUPHH 498
DB 1604 HLLT-----TLPOH 1612

RESULT 5
AE1810
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE1810
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1227 <KUR>
A:Cross-references: GB:BA0000019; PIDN:BAB77553.1; PID:g17135007; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0029

Query Match 16.4%; Score 444.5; DB 2; Length 1227;
Best Local Similarity 29.2%; Pred. No. 8.4e-24;
Matches 130; Conservative 73; Mismatches 171; Indels 71; Gaps 15;

QY 64 VSINEDG-TLFDGRPIESLSIDAVMPDVVQTRQ-----QAYRDKLAQQOAAAAA 117
DB 780 VSFSPDGQTLASGSDSTVLWD-----VKTQCQWQIFEGHKKVSVRFSPDQTLAS 833

QY 118 ASQOQSAK-----NGE--NTANGENGANTIANHNTDMMEVDG-----DV 155

DB 834 CGEDRSIKLWDIQGECVNTLWGHSSQVWAIA-----FSPDGRITLISCDQDQTLARLWDV 887
QY 156 EIPPNAKAVLVRGHESEVFCANVPVSDLLASGSGDSTARIWNLSENSTSGSTQVLVRHCI 215
DB 888 -ITGNSLNIIRGYTRDYSVAFSPDSQILASGDDYITGLWNLKTGEC-----HPL 937
QY 216 REGGQDVPSNKDVSTLDWNSGTTLLATSGSYDGFARIW-----TKDGNLSTLQCHKGPIFA 271
DB 938 R-GHQG-----RIRSVAFHPDGKILASGSADNTIKLWDISDTNHSKYIRLTGHTNWVWT 991
QY 272 LKNKKGNFILLSAGVDKTTIILDAHTGEAKQOFPFHSAPALVDWQ-SNNTFASCSTDMC 330
DB 992 WFSFDKHTTASSSEDRTIRLWDKDTGDCLOKLKSHHWVMTVAFSPDGRILASGSADSE 1051
QY 331 IHVCKLQDRIPIKTPQGHNTNEVNAIKWDPTGNLLASCSDMTLKIWSMKQDCNVHDLOQ 390
DB 1052 IKIWDVASGKCLQTLTDPQMIWSVAFSLDGTLLASASEDTVKLWNLKTGECVHTLK 1111
QY 391 NKEIYTIKWSFTGPGTNNPNANMLASASPDSTVRLWDVDRGICHTLTK-HQEPVVSFA 449
DB 1112 EKQVSVAFSPNGQ-----IAASGSEDTTVKLWDISTGSCVDTLKHGHTAAIRSA 1162
QY 450 FSPDGRYLAGSPDKCVHIWNTQVC 474
DB 1163 FSPDGRULASGSEDEKIQLWDMQNC 1187

RESULT 6
AI2493
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AI2493
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1189 <KUR>
A:Cross-references: GB:BA0000020; PIDN:BAB78213.1; PID:g17135667; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7129
A:Genome: plasmid

Query Match 15.8%; Score 429.5; DB 2; Length 1189;
Best Local Similarity 30.3%; Pred. No. 9.7e-23;
Matches 113; Conservative 63; Mismatches 134; Indels 63; Gaps 12;

QY 165 LRGHESEVFCANVPVSDLLASGSGDSTARIWNLSENSTSGSTQVLVRHCIREGGQDVPS 224
DB 726 LQGHQHWKTIAFNSGGRILASGSFDQNVKLWDIH-----TKCVMTL-----QGHGV-- 774

QY 225 NKDVSTLDWNSGTTLLATSGYDGFARIW-TKDGNLSTLQCHKGPFLAKWKKGNFILLS 283
DB 775 ---VTSVAFNPKDNLGSGSYDQSVKVDRTKRCCLDTLKKHTNRIWSVAFHPQGHILFVS 831

QY 284 AGVDKTTIILDAHTGEAKQOFPFHSAPALDV---DWQSNNTFASCSTDMCHVCKLG---- 337
DB 832 GGDHAAKIWEIQTGQCIKTFQGHSNATYIAHNWE-HSLLASGHEDOTIKLMDNLHSP 890

QY 338 -----QDRPIKTPQGHNTNEVNAIKWDPTGNLLASCSDMTLKIWSMKQDCNVHDLOQHNK 392
DB 891 HKSNNVTHPFRILQCHSNRVFSVVSFTGQLLASGSADRTIKLWSPTGQCLHLTHGHGS 950

QY 393 BIYTIKWSPTGPGTNNPNANMLASASPDSTVRLWDVDRGICHTLTKHQEPVVSFAFSP 452
DB 951 WWAIAFSLD-----DKLLASGSYDHTVKIWDYSSGQCLQTLQGHGSLVLAFAFC 1001

RESULT 8
Tl8521
beta transducin-like protein - Podospora anserina
C:Species: Podospora anserina
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl8521
R:Saupé, S.; Turcq, B.; Begueret, J.
Gene 462, 135-139, 1995
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
A:Reference number: Z18944; MUID:96009891; PMID:7557402
A:Accession: Tl8521
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1356 <SAU>

Query Match	15.0%;	Score 408;	DB 2;	Length 1711;	
Best Local Similarity	25.2%;	Pred. NO. 5.7e-21;			
Matches 125;	Conservative	76;	Mismatches 182;	Indels 94;	Gaps 15;
QY	65	SINEDGTL----	FDGRIEISL-----	LIDAVMPDVWQTCQAYRDKLAQOQAAAAA	114
Db	1246	SASDDGIRLWSLDGRLLIIPSHTKQVLAVTSPD-----	-----	-----	GGTIVSA 1288
QY	115	AAAAQQGSAKNGE--	NTANGEENG-----	-----	AHTIANHDTDMWVDGDEVEIPP 159
Db	1289	GADNTVKLWSRNGTLLTLEIGHNEAVWQVIFPDGRLIATASADKTIILWSRDGNI-----	-----	-----	1344
QY	160	NKAVVLRGHESEFICAWNPFVSDLLASGGDGDSTARIW-----	-----	-----	196

Db 1345 --LGTAGHNEVNSLSFSPDGNILASGSDNTVRLTWNRLPKTFYGHKGSVVRPS 1402
QY 197 NUGENSTSGTQVLVHCHTREG-----GQDVPSNKKVDTSLDWNSEGTLLATGSDVDFAR 250
Db 1403 NDGKKTISLSTSTMTWSLDGKLLQTLSPPLP--DVTISFTPDNKIVALASPDHTH 1459
QY 251 IWTKGNLASTLQHGKGPFAFKWKKGNFILLSAGVDKTTIWDHTGEAKQFPFHSAP 310
Db 1460 LYNROGLRLSPGHNHTISFSFNKQILASGSADKTIKLSVN-GRLLKTLIGHNGW 1518
QY 311 ALDVEDQSN-NTFASCTDMCHVCKLQGDPRIPKTFQHTNEVNAIKWDPTGNLLASCSD 369
Db 1519 VTDIKESADGNIVSADKTIKWSL-DGRLIRTLQGHSAVWSVNLSPDQTLASTSQ 1577
QY 370 DMTLKWSMKQDNCVHDLQHNKEIYIKWSPGTGPTNNPNANMLASASPDSTVRLWDY 429
Db 1578 DETIKLWNLNGE-LIYTLRGHSDVVNLSFSPDGK-----TIASASDGTIKLWNV 1627
QY 430 DRGICHTLTKHQEPYVSAFSPDGRYLASGDFDKCVHIWNTQ-VCLHYINGQVILN 485
Db 1628 PNGTLAKTFQHRGGVRSVSPDGKILASGGHDTTVKWNLEGIELQTLNLDLNL 1684
RESULT 10
AGI889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AGI889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; UID:21595285; PMID:11759840
A:Accession: AGI889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <R>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:g17130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664
Query Match 14.9%; Score 404.5; DB 2; Length 934;
Best Local Similarity 24.1%; Pred. No. 4.4e-21;
Matches 124; Conservative 104; Mismatches 176; Indels 111; Gaps 19;
QY 8 VNFVRYLQESGFSSAFTFGIKSHISQSNINGALVPPAALISIIQKLVVRAEVSIN 67
Db 231 ISSLVTLGLALGLAGVWLQGGKARMSETK-----AISSSAESFINANLEF- 277
QY 68 EDGTLFD---GRPIESLSLIDA-VMPDVVQTRQAY-----RDKLAQQQAAAAA 118
Db 278 -DGLIASIRARRIKGTGDIGANTRTQITETIQQSINFVREKNRLAEHGMLESVSPD 336
QY 119 SQGSNKGENTANGENGAHTIANHNTDMEVDGVEIPPNKAVLVLRGHESEVP-ICAW 177
Db 337 SXFIATASDKTVK-----IWSLDGKKQD-----VYLREEKEGENSVAF 376
QY 178 NPVSDLLASGSDSTARIWNLSNSTSGTQLVLRHCIREGQ--DVPSNKD-VTSLDWN 234
Db 377 SPDGTLMTGSDWNTAKWS-----REGRLHTLDGKHAIVLEAFS 418
QY 235 SEGTLTATGSDGFARITWKDGNLSTLQGHKGPFAFKWKKGNFILLSAGVDKTTIWD 294
Db 419 PDSQLLATASWNTVRLWSREGKLLHTLEGHKDKVNSITFSPDQGLIATVGDWNTMKLN 478
QY 295 AHTGEAKQFPFHSAPALVDWQ-----SNNTFASCTDMCHVCKLQGDRIKTFQHT 349
Db 479 L-DGKELTRFRHQ-----DMIVSVSFPDGKQIATASGDRTVKWSL-DGKELQTLRGH 532
QY 350 NEVNAIKWDPTGNLLASCSDMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPG---- 405

Db 533 NGVNSVTFSPDGKLIATASGDRITVKLWNSKQBP-LETLYGHTDAVNSVAFSPDGTSTATA 591
QY 406 -----TNNPNANML-----LASASFDSTVRLWDV--DRGICI 435
Db 592 GNDKTAIKWLNSPNSIIVRGHEDEVDLVFSPNGKYIATASWKTAKLWSIVGDKLQEL 651
QY 436 HTLKHQEPYVSAFSPDGRYLASGDFDKCVHIWN 470
Db 652 RTFNGHQRVKNKLSFSPDGKYIATTSWDKTAKLWN 686
RESULT 11
AC1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; UID:21595285; PMID:11759840
A:Accession: AC1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <R>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:g17135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283
Query Match 14.5%; Score 395; DB 2; Length 1747;
Best Local Similarity 28.2%; Pred. No. 5.1e-20;
Matches 105; Conservative 59; Mismatches 131; Indels 78; Gaps 9;
QY 165 LRGHSEVFICAMPVSDLLASGSDSTARINWL-----SENSTSGSTQ 208
Db 1310 LTGHERITSVKFSPDGKILASAGDKTIKFNTDGLKFLKTIAAHNQVNSINFSPDST 1369
QY 209 LV-----LRHCIREGGQDV-----PSNK----- 226
Db 1370 LVSGAGSTWKKWKIDGTILKTISGRGEQIRDVTFSPDNKVIASASDKTVIRQLNYQK 1429
QY 227 ---DVTSLDWNSEGTLLATGSDGFARITWKDGNLA---STLQGHKGPFAFKWKKKG 278
Db 1430 SOKSNVSVSPNPGKTFASAGWDGNITWQRE-TLAHSSLSITIKNQNIITVSYSPDG 1488
QY 279 NFILSAGVDKTTIWDHTGEAKQFPFHSAPALVDWQ--SNNTFASCTDMCHVCKLQ 337
Db 1489 KTIATASADNTIKLWDSQTQOLIKTLTGHKDRITTLTFHPDNQTIASGSADKTIKIRVN 1548
QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASGSDMTLKWSMKQDNCVHDLQHNKEIYTI 397
Db 1549 DGQLRLITLGHNDVTSVNFSPDQGLASGSDTNTVKIW-QTDGRLIKNITGHGLATSV 1607
QY 398 KWSPTGCTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQEPYVSAFSPDGRYL 457
Db 1608 KFSPD-----SHTLASASWNTIKLQWVTDGKLINNLNGHIDGVTSLSFSFDPGEIL 1658
QY 458 ASGSDFDKCVHIWN 470
Db 1659 ASGSADNTIKLWN 1671
RESULT 12
S76086
beta transducin-like protein, 190K - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sl0163
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000

C:Accession: S76086
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S.
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76086
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1693 <XAN>
 A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PID:BA010064.1; PID:dl01071
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:1051-1084/Domain: WD repeat homology <WD01>
 F:1092-1125/Domain: WD repeat homology <WD02>
 F:1133-1166/Domain: WD repeat homology <WD03>
 F:1174-1207/Domain: WD repeat homology <WD04>
 F:1256-1289/Domain: WD repeat homology <WD05>
 F:1297-1330/Domain: WD repeat homology <WD06>
 F:1338-1371/Domain: WD repeat homology <WD07>
 F:1420-1453/Domain: WD repeat homology <WD08>
 F:1461-1494/Domain: WD repeat homology <WD09>
 F:1502-1535/Domain: WD repeat homology <WD10>
 F:1584-1617/Domain: WD repeat homology <WD11>
 F:1625-1658/Domain: WD repeat homology <WD12>
 Query Match 14.5%; Score 394.5; DB 2; Length 1693;
 Best Local Similarity 28.2%; Pred. No. 5.3e-20;
 Matches 115; Conservative 55; Mismatches 149; Indels 89; Gaps 12;
 QY 100 RDLAQOQAAAAAASQGSAXNGENTANGE--ENGAHTIANHNDMEVDGVEI 157
 DB 1362 RDKTARLWTTTEGSCVAVLADHQGVREGQFSPDQGVITGS---ADKTAQLNVLG---- 1414
 QY 158 PPKAVVLRGHSEVFICANVPVSDLLASGSGSTARIWNLSNSGTSQVLVLRHCIRE 217
 DB 1415 --KKLTVLRGHQDVLNVRSPDSQVITVASKDGTARVN-----NTGRELAVLRH---- 1463
 QY 218 GGGVPSNKDVTSLDWNSEGTLIATSYGDFARIWTKDGNLSTLQKHGPTTFALKWNKK 277
 DB 1464 -----YEKNIFAEFSADQGFIVTASDDNTAGIWEIVRGVIGCRGHEGVPVFAQFSAD 1517
 QY 278 GNFTLSAGVDKTIINDAHTGEAKQFPFHSAPALVDQSNNTFASCTDMCIHVCKLG 337
 DB 1518 SRYLTASVDNTARIWDF-----LJ 1537
 QY 338 QDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTLKIMSKMKNODNCVHDLQOHKKEIYTI 397
 DB 1538 --RPLLTLAGHQSIYVQARFSPENGLIATVSADHTARLWD-RSGKTAVLYGHQGLVGTV 1594
 QY 398 KNSPTGCTNNPNANLMLASASFDSTVRLVDVRGICHTLTKHQPVSVAFPSDGRYL 457
 DB 1595 DWSPDQ-----MLVTASNDGTARLWDLG--GRELLTLEGHGNVRSASFSPDGRVW 1644
 QY 458 ASGSFDCVCHWNTQVCLHYLN--GOVLLNLRGSIICLYTLPHLVVIP 503
 DB 1645 LTSSADGTAKLPVKVTLPLQLSQGGWLKNY-----LTHNALVSP 1684
 RESULT 13
 T40883
 WD repeat protein - fission yeast (*Schizosaccharomyces pombe*)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T40883
 R:Wood, V.; Rajandream, M.A.; Barrrell, B.G.; Murphy, L.; Harris, D.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21954
 A:Accession: T40883
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-564 <WOO>

A:Cross-references: EMBL:AL031764; PIDN:CAA21113.1; GSPDB:GN00066; SPDB:SPCC1235.09
 A:Experimental source: strain 972h-; cosmid c1235
 C:Genetics:
 A:Gene: SPDB:SPCC1235.09
 A:Map position: 3
 A:Introns: 18/1; 273/3; 413/3

Query Match 14.1%; Score 382.5; DB 2; Length 564;
 Best Local Similarity 24.4%; Pred. No. 8.5e-20;
 Matches 131; Conservative 106; Mismatches 198; Indels 101; Gaps 24;

QY 3 ISSDEVFLVRYLQESGFSHSAFTFGIKSHISOSNIN---GALVPPAALISIIQKGLQY 59
 DB 1 MDTQVYIIWRYLKEGYSHTKFAFERETGI---QNLDKQWGSTCQVGAUVEILQKGLQY 58
 QY 60 VEAEV-----SINEDG--TLFDCRPISLSLID-----AVMPDVVQT----- 94
 DB 59 VELEKHYVDNHSNEBASKTSIDG-----SLVNNPCKLPFLVTVPHICETTLTKADST 113
 QY 95 -----RQAYRDKLAQQQAAAAAASQGSAXNG-----ENTANGENGHTIA 142
 DB 114 NGFCHEHNSNDHQILQDKGSGSPSPVMPFKDKIEKRDIDITMADESINVEKDPARPIA 173
 QY 143 NNHTDMMVEVDGVEIPNKAVLVRGHE---SEVF-----ICA-WNPVSD-----LLA 185
 DB 174 VNSSPV-----TETIKQVFTTGEDIKSDFFKVIPTKHPVTCADWRELLQENHYVE 228
 QY 186 SSGSDSTARIWNLS-----ENSTGSTQVLVLRHCIREGGQVPSNKDVTSLDWNSEGTLLA 241
 DB 229 FSIQMTNATLASVICEQNDFAKTD---YCL-----QSSFDNQDITGVAMNNSGFLA 280
 QY 242 TGSVDGPARIWKDGNLSTLQKHGPIFALKWNKKNFILSAGVDKTIINDAHTGEAK 301
 DB 281 YAFSGVIEIYDSHGSQLSFHNKKGVLKWSGTTIYLAAGSADGTITLFD-----QLK 336
 QY 302 Q-QPFPH--SAPALVDVQSNNTFASCTDMCIHVCKLQGDRIKTF-QGHTNEVNAIKW 357
 DB 337 QTVQSIDTLASSVLDIEWISDFEFTSDVEGSLRVYKVDGKAPVSTVSHADHNSIALRY 396
 QY 358 DPTGNLLASCSDDMTLKIMSKQD---NCVHDLQOHKKEIYTIKWSPTGPTGNPNANLM 414
 DB 397 NLRSLLLTASDSTVTKLWSRGDAGAFECLEH-VFSFSPVNCIDW-----NLRGTPI 448
 QY 415 LASASFSTVRLVDVRGICHTLTKHQPVSVAFPSDGRYLASGSGFDCVCHW 470
 DB 449 LAVAS-NSIVGMVMAISLQQLAVPMRHTAPVSALSFHNGRYLATGDTSGGVCWS 503

RESULT 14
 AB2410
 WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AB2410
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2410
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1551 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA076533.1; PID:g17133971; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4834

Query Match 14.0%; Score 381; DB 2; Length 1551;
 Best Local Similarity 23.4%; Pred. No. 4.4e-19;
 Matches 118; Conservative 94; Mismatches 175; Indels 118; Gaps 18;


```
QY 48 ALISIIKGLQYVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYDKLAQQ 107
Db 903 ALMSAURSG-----KALQALVKGRSLAKYPAISPLALOTILDNIQERNQ-----FQGHQ 953
QY 108 AAAAAAASQGSAXKNGENTANGENGHAHTIANHHTDMMEVDGVEIPPNKAVILRG 167
Db 954 AWRSVSFSRSGQYIILTASDDCTAR-----LWNLQ-----KQLISLQG 992
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTS-----GSTQLVL-----RHCIR 216
Db 993 HEDTIWSANFSPGKVIATASSDRTARLWNFSGQQLAKFQGHQYVRSVSPDGKHAT 1052
QY 217 EG-----GQDV-----PSNKDV-TSLD-----WNSEGTLL-- 240
Db 1053 AGDHTARLWFSFGQQLVQFPGHQGTVMCISFSPDGKHIATAADDRIVELWNLKGLLVR 1112
QY 241 -----ATGSDGPFARIWTKGNLSTGLQHKGPFPALKWKNKGNF 280
Db 1113 FPGHQCVMDVSPDSQVIATASSDGTSLRLNLAGEQITRFRGHQGVVSVRFSFNGQY 1172
QY 281 ILSAGVDKTTIIDAHTGEAKQFPFHSAPALDVDMQSNNTF-ASGSDTMCIHVCKLGOD 339
Db 1173 IATSSDRTARVWNLN-GQLAQFSGHQDYVRSVSPDGKIATASSDRTVRLWHLNKO 1231
QY 340 RPIKTFQGHNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCVHDLQQHNKEIYTIKW 399
Db 1232 Q-FSAFQGHQSTVRSVDFSPDGQKVTAADDRTVRLWNLKGBELLQFL-GHRGKVMVSF 1289
QY 400 SPTGPTNNPNANMLASAFSTVRLWDVDRGICHTLTKEHPVYSVAFSPDGRYLAS 459
Db 1290 SDGK-----YIATSSDRTVRLWDI-TGQLLQPPGHQGTVMVSFSPDGQHIAT 1339
QY 460 GSPDKCVHIWNTQVCLHYLNGQVLL 484
Db 1340 ASSDLTTRLWS-----LDGQELM 1357

RESULT 15
AH2195
hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2195
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; UID:21595285; PMID:11755840
A:Accession: AH2195
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074818.1; PID:gi17132214; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3119

Query Match 13.6%; Score 369; DB 2; Length 676;
Best Local Similarity 23.7%; Pred. No. 1e-18;
Matches 116; Conservative 96; Mismatches 194; Indels 84; Gaps 13;

QY 21 FSHSAFTGKISHISOSNINGALVPPAALISIIQGLQYVEA-----EVS 65
Db 226 FSLGATCFHLLTGINPSNL-----FVEQYSWVESWQQYWNSTNSDRNEGYL 273
QY 66 INEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYDKLAQQAAAAA-----AAAAAASQGG 122
Db 274 VKVLNKLLETDIQRYYQSADAEVWMDLTKQSLSLRKLTKTTPKSAIFRSWSASTSLTAST 333
QY 123 SAKNGENTANGENGHAHTIANHHTDMMEVDG-----DVEIPPKNKAVILRG 167
Db 334 TTKQAWKLNGRLK-QQLINTMSALLGLVGVGHLSQPLITKFSBISTQP---YTLKG 389
```

```
QY 168 HESEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLHRCIREGGQDVPSNKG 227
Db 390 HASDVNSVAFSPNGEFLASGSDDKTIKVMNLKN-----KQKIHTLPQHS 434
QY 228 -VTSLDWNSEGTLLATGSDGFARIWTKDGNLAS-----TLGQHKGPFPALKWKNKGNFI 281
Db 435 WWAIAFSPDGKTLASTGADXTIKLW-----NLATGKEIRHLKHSQGVASVAFSPDGKTL 490
QY 282 LSAGVDKTTIIDAHTGEAKQFPFHSAPALDVDMQ-SNNTFASCSCTDMCIHVCKLGODR 340
Db 491 ASGLDKTIKLNWPAATCKEIRTLQEHSSGVANVAFSPDGKTLASGSDKTIKLMNLTTSK 550
QY 341 PIKTFQGHNEVNAIKWDPGTGNLLASCDDMTLKIWSMKQDNCVHDLQQHNKEIYTIKW 400
Db 551 VIHLKGHSDLVMSVAFNSDSQTLASGSKTKIKLWNLSTGKTIRTLRGHSDKVNVAIV 610
QY 401 PTGPTNNPNANMLASAFSTVRLWDVDRGICHTLTKEHPVYSVAFSPDGRYLAS- 459
Db 611 -----PRDSTVLASGSDNTIKLWLTGELIRTKRDSGYIYSIVISPDGRLASG 662
QY 460 GSPDKCVHIW 469
Db 663 GSAENIIRKIW 672
```

Search completed: August 9, 2004, 16:48:49
Job time : 14.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:40:42 ; Search time 8.66667 Seconds
(without alignments)
3088.161 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 2718
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHHLVLPVALTELLVVK 514

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2493	91.7	514	1	TBLR_HUMAN
2	2229	82.0	526	1	TBLX_HUMAN
3	2170	79.8	522	1	TBLX_HUMAN
4	1308	48.1	313	1	TBLX_MOUSE
5	475	17.5	1526	1	YV46_ANASP
6	472	17.4	535	1	YF22_YEAST
7	458.5	16.9	1258	1	YS00_ANASP
8	451.5	16.6	1683	1	YL24_ANASP
9	420.5	15.5	1356	1	HET1_PODAN
10	394.5	14.5	1693	1	Y163_SYNY3
11	366.5	13.5	485	1	HU57_HUMAN
12	365.5	13.4	485	1	HU57_MOUSE
13	351.5	12.9	409	1	LI51_HUMAN
14	350.5	12.9	409	1	LI51_BOVIN
15	350	12.9	339	1	CI41_HUMAN
16	349.5	12.9	409	1	LI51_MOUSE
17	339	12.5	800	1	T2D4_HUMAN
18	337	12.4	515	1	YCW2_YEAST
19	333	12.3	742	1	PKA1_THECU
20	331.5	12.2	361	1	WDS3_DROME
21	331	12.2	437	1	AAC3_DICDI
22	325.5	12.0	334	1	WD85_HUMAN
23	319.5	11.8	606	1	PF20_CHLRE
24	317	11.7	579	1	SE10_CAEEL
25	310	11.4	605	1	POF1_SCHPO
26	309.5	11.4	376	1	YK44_CAEEL
27	308	11.3	704	1	T2D4_DROME
28	307	11.3	798	1	T2D4_YEAST
29	305	11.2	665	1	LI23_CAEEL
30	300	11.0	501	1	YH32_CAEEL
31	299	11.0	640	1	MT30_YEAST
32	291	10.7	1191	1	Y143_SYNY3
33	289.5	10.7	395	1	GBB5_HUMAN

34	288.5	10.6	596	1	TU12_SCHPO
35	287.5	10.6	395	1	GBB5_MOUSE
36	286	10.5	678	1	SCOB_EMBNI
37	285	10.5	605	1	FW1A_HUMAN
38	283	10.4	395	1	YZLL_CAEEL
39	283	10.4	542	1	FW1B_HUMAN
40	283	10.4	931	1	LUG_ARATH
41	282.5	10.4	682	1	TUP1_KLULA
42	281	10.3	554	1	PRP4_ARATH
43	280.5	10.3	380	1	GBB_MAIZE
44	280	10.3	614	1	TU11_SCHPO
45	279	10.3	522	1	PRP4_HUMAN

ALIGNMENTS

RESULT 1					
TBLR_HUMAN					
ID	TBLR_HUMAN	STANDARD;	PRT;	514	AA.
AC	Q9BZK7; Q9H211; Q9H9A1;				
DT	28-FEB-2003 (Rel. 41, Created)				
DI	28-FEB-2003 (Rel. 41, Last sequence update)				
DE	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Nuclear receptor co-repressor/HDAC3 complex subunit TBLR1 (TBL1-related protein 1).				
OS	TBLR1 OR IRA1.				
GN	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
[1]	SEQUENCE FROM N.A., AND INTERACTION WITH HDAC3.				
RP	MEDLINE=21930350; PubMed=11931768;				
RA	Zhang J., Kalkum M., Chait B.T., Rosser R.G.;				
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";				
EL	Mol. Cell 9:611-623(2002).				
[2]	SEQUENCE FROM N.A.				
RP	Zhang X., Dormady S., Basch R.;				
RA	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";				
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.				
[3]	SEQUENCE FROM N.A.				
RP	ISOgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y., Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;				
RT	"NEDO human cDNA sequencing project.";				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBUNIT: Interacts with HDAC3.				
CC	-1- SIMILARITY: Contains 1 Lish domain.				
CC	-1- SIMILARITY: Contains 8 WD repeats.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; AF14544; ARK00301.1; -				
DR	EMBL; AF268193; AAC44736.1; -				
DR	EMBL; AK022956; BAB14331.1; -				
DR	InterPro; IPR006594; Lish.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF00400; WD40; 8.				
DR	PRINTS; PR00320; GPROTEINRPT.				
DR	ProDom; PD000018; WD40; 3.				
DR	SMART; SM00657; Lish; 1.				

```

DR SWART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 167 206 WD 1.
FT REPEAT 223 262 WD 2.
FT REPEAT 264 303 WD 3.
FT REPEAT 306 344 WD 4.
FT REPEAT 347 386 WD 5.
FT REPEAT 389 437 WD 6.
FT REPEAT 440 479 WD 7.
FT REPEAT 481 513 WD 8.
FT DOMAIN 108 118 POLY-ALA.
FT CONFLICT 31 31 E -> K (IN REF. 2).
FT CONFLICT 59 59 Y -> H (IN REF. 3).
FT CONFLICT 389 389 A -> Q (IN REF. 2).
SQ SEQUENCE 514 AA; 55595 MW; 0B55662DEE4BA796D CRC64;

Query Match 91.7%; Score 2493; DB 1; Length 514;
Best Local Similarity 99.6%; Pred. No. 2.4e-174;
Matches 470; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISSEVFLVYRYLQSGFHSFSAFTFGIKHSISOSNINGALVPPAALISIIQKGLQV 60
DB 1 MSISSEVFLVYRYLQSGFHSFSAFTFGIESHSISOSNINGALVPPAALISIIQKGLQV 60

QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDVKLAQOAAAAAASQ 120
DB 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDVKLAQOAAAAAASQ 120

QY 121 QGSAKNGENTANGAHTIANNHDDMEVDGVEIPENKAVLGRGHESEVFIKANPV 180
DB 121 QGSAKNGENTANGAHTIANNHDDMEVDGVEIPENKAVLGRGHESEVFIKANPV 180

QY 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGQDVPSKDVTSLDWNSGTL 240
DB 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGQDVPSKDVTSLDWNSGTL 240

QY 241 ATGSGDGFARIWTKDGNLSTLGHQGPFLAKWKKGNFILLSAGVDKTIWDAHTGEA 300
DB 241 ATGSGDGFARIWTKDGNLSTLGHQGPFLAKWKKGNFILLSAGVDKTIWDAHTGEA 300

QY 301 KQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360
DB 301 KQGFPHSAPALVDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPGTNNPNNMLASASF 420
DB 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHNKEIYTIKWSPTGPGTNNPNNMLASASF 420

QY 421 DSTVRLWVDVDRGCIHTLTKHQBPVSVAFSPDGRVYLASGSPDKCVHIWNQ 472
DB 421 DSTVRLWVDVDRGCIHTLTKHQBPVSVAFSPDGRVYLASGSPDKCVHIWNQ 472

RESULT 2
TBX HUMAN STANDARD; PRT; 526 AA.
AC 060907;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transducin beta-like 1x protein (Transducin-beta-like 1, X-linked).
GN TBX1X OR TBX1.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99264241; PubMed=10330347;
RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,
RT "X-linked late-onset sensorineural deafness caused by a deletion
RT involving OAI and a novel gene containing WD-40 repeats.";
RL Am. J. Hum. Genet. 64:1604-1616(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph.
RX MEDLINE=2386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.
CC -----
CC EMBL; Y12781; CAA73319.1; ALT_INIT.
CC EMBL; BC032708; AAB32708.1; -.
CC Genew; HGNC:11585; TBLX.
CC MIM; 300196; -.
CC GO; GO:0007605; P:hearing; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC GO; GO:0007601; P:vision; TAS.
CC InterPro; IPR006594; Lish.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 8.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD000018; WD40; 3.
CC SMART; SM00667; Lish; 1.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00896; LISH; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS00082; WD_REPEATS_2; 6.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 POLY-ALA.
SQ SEQUENCE 526 AA; 57046 MW; 98922F88EC42F6E9 CRC64;

```

```
Query Match      82.0%; Score 2229; DB 1; Length 526;
Best Local Similarity 85.4%; Pred. No. 4.4e-155;
Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQV 60
DB 1 MSISDEVNLFVRYLQESGFSAFTFGIESHSOSNINGTLVPPAALISIIQKGLQV 60
QY 61 EAVSINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQQAAAAA----- 117
DB 61 EABISINEDGTPDGRPIESLSLIDAVMPDVVQTRQOAFREKLAQQASAAAAAATAA 120
QY 118 -----ASOQGSAGKNGENTANGENGATIANHNTDMVEVDGVEIPNKAIVLRG 167
DB 121 ATAAITTSAGVSHQNFSEKREATVNGEENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 179
QY 168 HSEVFICAWNPSVSDLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKG 227
DB 180 HSEVFICAWNPSVSDLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGHVDPSNKG 239
QY 228 VTSLDVNSGTLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILLSAGVD 287
DB 240 VTSLDVNTGTLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILLSAGVD 299
QY 288 KTTIWDATHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQG 347
DB 300 KTTIWDATHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQG 359
QY 348 HTNEVNAIKWDPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKWSPTGPGNN 407
DB 360 HTNEVNAIKWDPSGMLLASCDSDMTLKIWSMKQEVCIHDLQAHKEIYTIKWSPTGPTS 419
QY 408 NPANMLASAFSDSTVRLWDVDRG:CIHTLKHQEPVYVAFSPDGRVLAGSGSPDKCVH 467
DB 420 NPNSNMLASAFSDSTVRLWDVDRG:CIHTLKHQEPVYVAFSPDGRVLAGSGSPDKCVH 479
QY 468 INWNTQ 472
DB 480 INWNTQ 484

RESULT 3
TELY_HUMAN
ID TELY_HUMAN STANDARD; PRT; 522 AA.
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE "The DNA sequence of the human Y chromosome."
DE Transducin beta-like 1 protein (transducin-beta-like 1, Y-linked).
GN TELY OR TELY.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RT Wilson R.K., Waterston R.H., Page D.C.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF332220; AAK13472.1; -.
DR EMBL; AF332221; AAK13473.1; -.

```

```
DR EMBL; AF332222; AAK13474.1; -.
DR Genew; HGNC:18502; TBLIY.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT DOMAIN 4 36 LISH.
FT REPEAT 177 216 WD 1.
FT REPEAT 233 272 WD 2.
FT REPEAT 274 313 WD 3.
FT REPEAT 316 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 447 WD 6.
FT REPEAT 450 489 WD 7.
FT REPEAT 491 521 WD 8.
SQ SEQUENCE 522 AA; 56688 MW; 4E020216422442D8 CRC64;

Query Match      79.8%; Score 2170; DB 1; Length 522;
Best Local Similarity 84.1%; Pred. No. 8.7e-151;
Matches 406; Conservative 32; Mismatches 33; Indels 12; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQV 60
DB 1 MSISDEVNLFVRYLQESGFSAFTFGIESHSOSNINGTLVPPAALISIIQKGLQV 60
QY 61 EAVSINEGTLDPGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQQAAAAA----- 115
DB 61 EAEISINKDGFVDSRPIESLSLIVAVIPDVVQMQRAFGEKLTQQQASAAATEASAMAK 120
QY 116 -----AAASQGSAGKNGENTANGENGATIANHNTDMVEVDGVEIPNKAIVLRGHE 169
DB 121 AATWTPAIISQNPFPKREATVNGEENGAEHI-NHNSKPMIEDGVEIPNKAIVLRGHE 179
QY 170 SEVFICAWNPSVSDLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKGDT 229
DB 180 SEVFICAWNPSVSDLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGHVDPSNKGDT 239
QY 230 SLDNVNSGTLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILLSAGVDKT 289
DB 240 SLDNVNSGTLATGSGYGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILLSAGVDKT 299
QY 290 TIIWDAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGT 349
DB 300 TIIWDAHTGEAKQOFPFHSAPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGT 359
QY 350 NEVNAIKWDPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKWSPTGPGNNP 409
DB 360 NEVNAIKWDPSGMLLASCDSDMTLKIWSMKQACVHDLQAHKEIYTIKWSPTGPTS 419
QY 410 NANMLASAFSDSTVRLWDVDRG:CIHTLKHQEPVYVAFSPDGRVLAGSGSPDKCVH 469
DB 420 NSSIMLASAFSDSTVRLWDVDRG:CIHTLKHQEPVYVAFSPDGRVLAGSGSPDKCVH 479
QY 470 NTQ 472
DB 480 NTQ 482

RESULT 4
TELY_MOUSE
ID TELY_MOUSE STANDARD; PRT; 313 AA.
AC Q9QX67;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

Transducin beta-like IX protein (Fragment).
 TBL1X OR TBL1.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=129/SVJ;
 RA Botcherby M.R.M., Straw R., Clarke D., Greyststrong J.S., Weston P.,
 RA Hunter G., Kimberley C., Rhodes M.;
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RL
 CC -!- SIMILARITY: Contains at least 5 WD repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; F38006; CAB61534.1; -;
 CC MGD; MGI:1336172; Tbl1x.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 5.
 CC PRINTS; PR00320; GPROTEINERPT.
 CC ProDom; PD000018; WD40; 2.
 CC SMART; SM00320; WD40; 5.
 CC PROSITE; PS00678; WD_REPEATS_1; 3.
 CC PROSITE; PS00882; WD_REPEATS_2; 4.
 CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
 CC Repeat; WD repeat.
 CC NON TER 1 1 WD 1.
 CC REPEAT <1 24 WD 2.
 CC REPEAT 76 115 WD 3.
 CC REPEAT 160 198 WD 4.
 CC REPEAT 201 249 WD 4.
 CC REPEAT 252 291 WD 5.
 CC NON TER 313 313
 CC SEQUENCE 313 AA; 34808 MW; 3BC48683432DFB7 CRC64;
 Query Match 48.1%; Score 1308; DB 1; Length 313;
 Best Local Similarity 83.2%; Pred. No. 3.2e-88;
 Matches 243; Conservative 16; Mismatches 19; Indels 14; Gaps 4;
 QY 188 SGSTARIWNLSENSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSG- - - - - TLLAT 242
 DB 2 SGSTARIWNLSENSTSGTQVLVLRHCIREGGQDVPSNKNVTSLDWNSGHCQVPMVL 61
 QY 243 GSYDGFARITWKGNLASTLQGHKGPFPALKNWKNKGNFILSAGVDKTTIWDHTGEAKQ 302
 DB 62 QEYV- - - - - QKSNLASTLQGHKGPFPALKNWKNKGNFILSAGVDKTTIWDHTGEAKQ 114
 QY 303 QFPFHS-APALVDVQSNNTFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPTG 361
 DB 115 QFPFHSAPALVDVQSNNTFASCTDMCHVCKLGDRIKTFQGHTEVNAIKWDPTG 173
 QY 362 NLLASCSDDMTLTKWSMKQNCVHDLOQHNKEIYTIKWSPTGPTNPNANMLASASFD 421
 DB 174 MLLASCSDDMTLTKWSMKQNCVHDLOQHNKEIYTIKWSPTGPTNPNANMLASASFD 233
 QY 422 STVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKCVHWTQV 473
 DB 234 STVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKCVHWTQV 285
 RESULT 5
 Y46 ANASP STANDARD; PRT; 1526 AA.
 ID Y46 ANASP
 AC Q8YR11;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical WD-repeat protein alr3466.
 GN ALR3466.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RA "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120".
 RL DNA Res. 8:205-213(2001).
 CC -!- SIMILARITY: Contains 16 WD repeats.
 CC -!- SIMILARITY: Contains 1 pentapeptide repeat domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP003593; BAB75165.1; -;
 CC FIR; AC2239; AC2239.
 CC InterPro; IPR007111; NACHT NTPase.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 15.
 CC PRINTS; PR00320; GPROTEINERPT.
 CC ProDom; PD000018; WD40; 8.
 CC SMART; SM00320; WD40; 15.
 CC PROSITE; PS00678; WD_REPEATS_1; 11.
 CC PROSITE; PS00882; WD_REPEATS_2; 15.
 CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
 CC PROSITE; PS0837; NACHT; UNKNOWN 1.
 CC Hypothetical protein; Repeat; WD repeat; Complete proteome.
 KW REPEAT 334 376
 FT REPEAT 823 862 PENTAPEPTIDE.
 FT DOMAIN 823 901 WD 2.
 FT REPEAT 862 901 WD 3.
 FT REPEAT 904 945 WD 4.
 FT REPEAT 946 985 WD 5.
 FT REPEAT 988 1027 WD 6.
 FT REPEAT 1030 1069 WD 7.
 FT REPEAT 1072 1111 WD 8.
 FT REPEAT 1114 1153 WD 9.
 FT REPEAT 1156 1195 WD 10.
 FT REPEAT 1198 1237 WD 11.
 FT REPEAT 1240 1279 WD 12.
 FT REPEAT 1282 1321 WD 13.
 FT REPEAT 1324 1363 WD 14.
 FT REPEAT 1366 1405 WD 15.
 FT REPEAT 1408 1447 WD 16.
 FT REPEAT 1450 1491 WD 16.
 FT SEQUENCE 1526 AA; 170376 MW; D5DF272509BA4738 CRC64;
 Query Match 17.5%; Score 475; DB 1; Length 1526;
 Best Local Similarity 31.0%; Pred. No. 1.5e-26;
 Matches 119; Conservative 63; Mismatches 138; Indels 64; Gaps 10;
 QY 136 NGAHTIANHTDMVEVDGVEIPNKAV-VLRGHESEVFICAWNPVSDLLASGSDSTAR 194
 DB 1127 NGV-TLANGSDQIVRLWD- - - - - ISSKCLYTLQGHNTNNAVAFSPDGTATLASGSDQTVR 1183
 QY 195 IWLNS- - - - - TSGSTQLVLRHCIREGGQDVPSNKNVTSLDWNSGFTLLATGSDYDFAR 250
 DB 1184 LWDISSKCLYILQHTSW- - - - - VNSVFNPDGSLASGSDQTVR 1225
 QY 251 IW-TKGNLASTLQGHKGPFPALKNWKNKGNFILSAGVDKTTIWDHTGEAKQFPFHS 309

```

Dd 1226 LWEINSSKCLTFCGHTSWNSVFNPDGSLAGSGDKVRLWLDISSKCLHTFOGHT- 1284
Qy 310 PALVDVQSNNTF-----ASCSTDMCIHVCKLQDPRPIKTFQGHTEYNEVNAIKWDPGTGN 362
Dd 1285 -----NWVNSVAENPDGSLAGSGDQTVRLWEISSKCLHTFCGHTSWNSVTFSPDGT 1339
Qy 363 LLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPGTNNPNANMLASAFDS 422
Dd 1340 MLASGSDQDQVRLMWSISSGCLYTLFHTNWNVGSVIFSPDG-----AILASGSDQ 1390
Qy 423 TVRLWVDVRCIHTLTKHQPVPYVSAFSPDGLYASGSPDKCVHWNTOVCLHYLNGQV 482
Dd 1391 TVRLWSTSSGCLYTLQGHNNWGSIVFSPDGLTLLASGSDQDQVRLW----- 1437
Qy 483 LLNLGRSICLYTLPHLWVPLVA 506
Dd 1438 --NISGECYTLRHINSVRSVA 1459

RESULT 6
SIF2 YEAST
ID SIF2 YEAST STANDARD; PRT; 535 AA.
AC P38262;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SIF2-interacting protein SIF2.
GN SIF2 OR EMB1 OR YBR103W OR YBR0832.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9288C;
RX MEDLINE=95208357; PubMed=7900426;
RA Mannhaupt G., Stucka R., Ehrlie S., Vetter I., Feldmann H.;
RT "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL Yeast 10:1363-1381(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98315485; PubMed=9651685;
RA Cockell M., Renauld H., Watt P., Gasser S.M.;
RT "Sif2p interacts with Sir4p amino-terminal domain and antagonizes
telomeric silencing in yeast.";
RL Curr. Biol. 8:787-790(1998).
RN [3]
RP IDENTIFICATION IN A COMPLEX WITH HOS2; HST1; SNT1; CPRI; YIL112W AND
SET3.
RX MEDLINE=21567937; PubMed=11711434;
RA Fijnappel W.W.M.P., Schaft D., Roguev A., Shevchenko A., Tekotte H.,
RA Wilk M., Rigaut G., Seraphin B., Aasland R., Stewart A.F.;
RT "The S. cerevisiae SET3 complex includes two histone deacetylases,
Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation
gene program.";
RL Genes Dev. 15:2991-3004(2001).
CC -!- FUNCTION: Antagonizes telomeric silencing in yeast. May recruit
CC SIR4 to non-telomeric sites or repression.
CC -!- SUBUNIT: Interacts with SIR4 amino-terminal domain. Interacts with
CC a complex composed of SIN3 and RPD3. Identified in the Set3C
CC complex with HOS2, HST1, SNT1, CPRI, HOS4/YIL112W and SET3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; X78993; CAAS5606.1; --
DR EMBL; Z35972; CAAS5058.1; --
DR PIR; S48266; S48268.
DR GenOnline; 138646; --
DR SGD; S000307; SIF2.
DR GO; GO:0000118; C:histone deacetylase complex; IPI.
DR GO; GO:0005634; C:nucleus; IPI.
DR GO; GO:0017136; F:NAD-dependent histone deacetylase activity; IDA.
DR GO; GO:0045129; F:NAD-independent histone deacetylase activity; IDA.
DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.
DR GO; GO:0016575; P:histone deacetylation; IDA.
DR GO; GO:0045833; P:negative regulation of meiosis; IPI.
DR GO; GO:0000122; P:negative regulation of transcription from P...; IPI.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 5.
DR ProDom; PD000018; WD40; 1.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; FALSE_NEG.
DR PROSITE; PS0082; WD REPEATS 2; 4.
DR PROSITE; PS0294; WD REPEATS_REGION; 1.
KW Repeat; WD repeat; Chromatin_regulator; Nuclear protein.
FT DOMAIN 4 36 LISH.
FT REPEAT 163 208 WD 1.
FT REPEAT 218 257 WD 2.
FT REPEAT 259 298 WD 3.
FT REPEAT 315 354 WD 4.
FT REPEAT 357 396 WD 5.
FT REPEAT 399 437 WD 6.
FT REPEAT 440 479 WD 7.
FT CONFLICT 396 396 S -> C (IN REF. 1; CAAS5058).
SQ SEQUENCE 535 AA; 59145 MW; 25DD19AB2BF84B07 CRC64;

Query Match 17.4%; Score 472; DB 1; Length 535;
Best Local Similarity 25.8%; Pred. No. 6.2e-27;
Matches 140; Conservative 91; Mismatches 230; Indels 82; Gaps 15;

Qy 1 MSISDEVNLFVRYLQSGFSHSA-----FTFGIKSHISOSNINGALVPPAAL 49
Dd 1 MSITSEELNYLWRYCQEMGHEVSALALQDTRVLEDFEYKEHI-----PLGTL 50
Qy 50 ISITQKGLQVVEAEVSNEDG---TLFDGRIEISLIDAWPDVWQTRQAYDKLAQQ 106
Dd 51 VNLVQRGILYTESELMVDSKDISALNEHLSDFNLVQALQID-----KEKPEI 101
Qy 107 QAAAAAASOQSAKNGENTANGEAHTIANNHTDMMEVDGDEIIPPKNKKN 166
Dd 102 SSEGRFTLETNSKAGEDCASTVERETQDDTNSIDSSD--DLDFGVKI--LKEIV-- 155
Qy 167 GHESEVFCANPVSVD--LLASGSDSTARIWNLSENSTSGSQ-----LVLRCIREGG 219
Dd 156 -KLDNISSVTWNPLDESILAYGEKNSVARLARIIVETDQEGKKYWKLTIIAELRPFALSA 214
Qy 220 QDVPSNKDVTSLDWNSEGTLLATSGYDGFARIMTKGNLASTLQGHKGPFAKWNKKN 279
Dd 215 SSGKTTNQVTCIAWSHDGNSIVTGVENGELRNKTKGALLNVLNFRHAPIVSVKWKDGT 274
Qy 280 FILSAGYDKTIIIDWANTGEAKQFPF-----HSAP---ALDVMQSNNTFAS 324
Dd 275 HIISMDYENVITLWNVISGTMQHFELKETGGSSINAENSHSGDGLGVDEYVDDDFKI 334
Qy 325 CSTDMCIHVCKLQDPRPIKTFQGHTEYNEVNAIKWDPGTNLLASGSDDDTLKIWSMKQNCV 384
Dd 335 PGPKGALFVQITEKTTGKLGIGHGPISVLEFNDTNKLLSASDDGTLIHWGGNGNSQ 394
Qy 385 HDLQQHNKEIYTIKWSPTGPGTNNPNANMLASAFDSVRLWVDVRCIHTLTKHQP 444
Dd 395 NSFYGHSSQSVASWV-----GDDKVISCSMDGSLVRLSLKQNTLLALSIVDGV 444
Qy 445 VYSAFSPDGRYLASGSGFDKCVHWNTOVCLHYLNGQVLLNLG-RSICLYTLPHLWV 503

```



```

FT REPEAT 1153 1192 WD 3.
FT REPEAT 1195 1236 WD 4.
FT REPEAT 1280 1318 WD 5.
FT REPEAT 1320 1359 WD 6.
FT REPEAT 1361 1400 WD 7.
FT REPEAT 1402 1441 WD 8.
FT REPEAT 1444 1483 WD 9.
FT REPEAT 1486 1525 WD 10.
FT REPEAT 1528 1567 WD 11.
FT REPEAT 1570 1609 WD 12.
FT REPEAT 1611 1650 WD 13.
SQ SEQUENCE 1683 AA; 185459 MW; F7CB361FF54F7137 CRC64;

Query Match
Best Local Similarity 16.68; Score 451.5; DB 1; Length 1683;
Matches 123; Conservative 75; Mismatches 167; Indels 73; Gaps 11;

QY 64 VSINEDG-TLFDGRPIESLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQOG 122
Db 1245 LSPSPDGKTLASADKTIKLRADGKLVKT-LKGHNDVWDVNFSSDGAIASARDN 1303
QY 123 SAKNGENTANGEAHTIANNHTDMVEVDGVEIPNKAVILRGHESEVFICAWNPVSD 182
Db 1304 TIK-----LWNRHGIELE-----TFTGSGGYAVNFPDLSN 1335
QY 183 LLASGSDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGLLAT 242
Db 1336 IIASLSLNTIRLWQPLISPL-----EVLNAGSGYAVNFPDLSN 1380
QY 243 GSYDGFARW-TKGNLSTLQGHKGPFPALKWKKGNFILSAGVDKTTIWDHAHTGEAK 301
Db 1381 AGADGNIQLWHSQDGLLTLPGNKA-IVGISFTQGDILASANADTKVIRVRDGKAL 1439
QY 302 QQPPFHSAPALVDWQ-SNNTRASCSTDMCIHVCKLGQDRPKTFPGHTNEVNAIKWDP 360
Db 1440 KTLIGHDNEVKNVFPDQKTLASARDNTVKLWNSDGFKTKLKGHTDEVFVWSFSD 1499
QY 361 GNLLASCSDDMTLKWSMKQDNCVHDLQQRNKEIYTIKWSPTGPGTNNPNANMLASASF 420
Db 1500 GKIITASADKTIKLRDLSFSGNLKSLPAHNDLVSNVFNPDGS-----MLASTSA 1550
QY 421 DSTVRLWDVDRGICHTLTKHQPVPYVAFSPDGRVLASGSDKCVHIWNTQVCLHYLNG 480
Db 1551 DKTVKLWRSHDGHLLTFSGHNVVSYSPDGRYIASASEDKTKVQW-----IDG 1603
QY 481 QVLLNLGRSICLYTLPHH 498
Db 1604 HLLT-----TLPOH 1612

```

RESULT 9

```

HET1_PODAN
ID HET1_PODAN STANDARD; PRT; 1356 AA.
AC Q00808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative incompatibility protein HET-E-1.
GN HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]_
SQ SEQUENCE FROM N.A.
RP MEDLINE=96009891; PubMed=7557402;
RA Saue S., Turcq B., Begueret J.;
RT "A gene responsible for vegetative incompatibility in the fungus
RT Podospora anserina encodes a protein with a GTP-binding motif and G
RT beta homologous domain."
RL Gene 162:135-139(1995).
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,

```

```

CC HET-C.
CC -!- SIMILARITY: Contains 10 WD repeats.
CC -!- SIMILARITY: Contains 1 NACHT domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L28125; AAA85775.1; -
DR PIR; T18521; T18521.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00578; WD REPEATS_1; 10.
DR PROSITE; PS0082; WD REPEATS_2; 10.
DR PROSITE; PS00294; WD-REPEATS_REGION; 1.
KW GTP-binding; Repeat; WD repeat.
FT DOMAIN 294 629 NACHT.
FT NP_BIND 300 307 GTP (POTENTIAL).
FT REPEAT 839 869 WD 1.
FT REPEAT 881 911 WD 2.
FT REPEAT 923 953 WD 3.
FT REPEAT 965 995 WD 4.
FT REPEAT 1007 1037 WD 5.
FT REPEAT 1049 1079 WD 6.
FT REPEAT 1091 1121 WD 7.
FT REPEAT 1133 1163 WD 8.
FT REPEAT 1175 1205 WD 9.
FT REPEAT 1217 1247 WD 10.
SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match
Best Local Similarity 15.5%; Score 420.5; DB 1; Length 1356;
Matches 117; Conservative 63; Mismatches 164; Indels 71; Gaps 12;

QY 78 IESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQOGSAKNGE-----NT 130
Db 820 ISTISVVEAEWNACTQT-----LEHGSSVLVSFAFGQVRSAGSDDKTIKIWD 870
QY 131 ANGENGAHTIANNHTDMVEVDGVEIPNKAVILRGHESEVFICAWNPVSDLLASGSD 190
Db 871 ASG--TGTQT-----LEHGSGVMSVAFSPDRERVASGSD 904
QY 191 STARIWNLSENSTSGS-TQLVLRHCIREGGQDVPSNKDVTSLDWNSEGLLATGSGYGA 249
Db 905 KTIKIW-----DAASGTCTQTLGEG-----VQSVAFSPDQGVRSAGSDDHTI 949
QY 250 RIW-TKDNGLASTLQGHKGPFPALKWKKGNFILSAGVDKTTIWDHAHTGEAKQPPFHS 308
Db 950 KIWDAAASGTCTQTLGEGSSVLVSFAFGQVRSAGSDDKTIKIWDASGTCTQTLGEG 1009
QY 309 APALVDWQ-SNNTFASCSTDMCIHVCKLGQDRPKTFPGHTNEVNAIKWDPGNLASC 367
Db 1010 GSVMSVAFSPDQGVRSAGSDDKTIKIWDASGTCTQTLGEGGVQSVQSVFSPDQGVRSAG 1069
QY 368 SDDMTLKWSMKQDNCVHDLQQRNKEIYTIKWSPTGPGTNNPNANMLASASPDSTVRLW 427
Db 1070 SDDHTIKIWDVAGSTCTQTLGEGSDSVMSVAFSPDQ-----RVAGSGIDGTIKIW 1120
QY 428 DVDRGICHTLTKHQPVPYVAFSPDGRVLASGSDKCVHIWNTQVCLHYLNG 480
Db 1121 DAASGTCTQTLGEGGVHVSFAFGQVRSAGSDDKTIKIWDAAASGTCTQTLGEG 1175

```

RESULT 10
Y163 SYNY3 STANDARD; PRT; 1693 AA.
AC Q55563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein sil0163.
GN SIL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64 to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- SIMILARITY: Contains 16 WD repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D63999; BAA10064.1; -;
DR PIR; S76086; S76086.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 16.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 13.
DR SMART; SM00320; WD40; 16.
DR PROSITE; PS00678; WD REPEATS 1; 8.
DR PROSITE; PS50082; WD REPEATS 2; 15.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042 WD 1.
FT REPEAT 1053 1083 WD 2.
FT REPEAT 1094 1124 WD 3.
FT REPEAT 1135 1165 WD 4.
FT REPEAT 1176 1206 WD 5.
FT REPEAT 1217 1247 WD 6.
FT REPEAT 1258 1288 WD 7.
FT REPEAT 1299 1329 WD 8.
FT REPEAT 1340 1370 WD 9.
FT REPEAT 1381 1411 WD 10.
FT REPEAT 1422 1452 WD 11.
FT REPEAT 1463 1493 WD 12.
FT REPEAT 1504 1534 WD 13.
FT REPEAT 1545 1575 WD 14.
FT REPEAT 1586 1616 WD 15.
FT REPEAT 1627 1657 WD 16.
SQ SEQUENCE 1693 AA; 189935 MW; 0977A827A0251CFF CRC64;
Query Match 14.5%; Score 394.5; DB 1; Length 1693;
Best Local Similarity 28.2%; Pred. No. 1.3e-20;
Matches 115; Conservative 55; Mismatches 149; Indels 89; Gaps 12;
QY 100 RDKLAQQQAAAAAASQSSAKKNGENTANGE--ENGATHTIANNHTDMMEVDGDVEI 157
DB 1362 RDKTARLTWTTEGECVAVLACHQGVNREGQSPDQGVITGS---AKTQALNNVLG---- 1414
QY 158 PPNKAVVLRGHESEVFCANPNPISDLASGSGSTARIWNLSENSTSGTQVLVLRHCIRE 217
DB 1415 --KKLTVLGRHQDVAIVNRFSPDSQYIVTASDKGTARVWN-----NTGRELAVALRH---- 1463

QY 218 GQDVPFNKDVTSIDMNSBEGTLLATGSDYGFARITWKDGNLASTLQGHKGPFPALKWNKK 277
DB 1464 -----YEKNIFAAPFSAQGFIVTASDDNTAGIWEIVGREGVIGRGHGPVYFAQFSAD 1517
QY 278 GNFILSAGVYKTIIDAHATGEAKQOPFPHSAPALDWDQSNNTFASCSTDCIHHVCKLG 337
DB 1518 SRYILTASVONTARIWDF-----LG 1537
QY 338 QDRPIKTFQGHTEVNAIKWDPGTGNLIASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTI 397
DB 1538 --RPLTLTLAGHSIVYQARPSPEGNLIATVADHTARLWD-RSGKTAVLVYGHQGLGVIV 1594
QY 398 KWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICIIHTLKHQEPYVSVAFSPDGRVL 457
DB 1595 DWSPDGQ-----MLVTASNDGTARLWDLG-SRELLTLEGHGNVRSABSFSPDGRVW 1644
QY 458 AGSFDKCVHIWNTQVCLHYLN--GOVLLNLGRSICLYTIPHLVVP 503
DB 1645 LTSSADGTAKLPVKVLPQLLSQGGQWLKXY-----LTHNALVSP 1684
RESULT 11
HUS7 HUMAN
ID HUS7 HUMAN STANDARD; PRT; 485 AA.
AC Q9NVX2; O60868; Q9BU54;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Nimomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]_SEQUENCE FROM N.A.
RC TISSUE=Lung, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
EL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]_SEQUENCE OF 400-485 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,

```
RA Cannata N., Zimbello R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RL yeast sequences.";
RN Yeast 18:69-80(2001).
RN [4]
RX SUBCELLULAR LOCATION.
RA MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001320; BA91621.1; -
DR EMBL; BC002884; AAH02884.2; -
DR EMBL; BC012075; AAH12075.1; -
DR EMBL; AJ005257; CRA06444.1; -
DR SWISS-2DPAGE; Q9NVX2; HUMAN.
DR InterPro; IPR001632; Gpotein_B.
DR Pfam; PF00400; WD40.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 7.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 325 366
FT REPEAT 370 409
FT REPEAT 412 451
FT REPEAT 454 484
SQ SEQUENCE 485 AA; 53266 MW; 0ABE24B44957379 CRC64;
Query Match
Best Local Similarity 13.5%; Score 366.5; DB 1; Length 485;
Matches 103; Conservative 47; Mismatches 145; Indels 89; Gaps 9;
QY 165 LRGHSEVFICAWNPVSDLLASGSDSTARIWNLSEN---STGSTQLVLRHCIREGGQ 220
DB 110 LECHSEAVISVAFSPGKYLASGSDTTRVFDLSTETPHFTCKGHRHWLISLSPDGK 169
QY 221 DYPS-----NKDVTSIDW-----NSEGITLLATGSDVDGAR 250
DB 170 KLASGCKGQILLDPSTGKQVRLTAGHSKWITLGSWEPLHANPECRYVASSSKDGSVR 229
QY 251 IW-TKDGNLASTIGQHKGP IFALKVKNKGNFILTASGVDKTTIWDHTGEAKQPPFH-- 307
DB 230 IWDTTAGRCERLTGHTQSVTCRLRGDGL-LIYSASQDRTIKVWHAHGVLCRTLQHGCH 288
QY 308 --SAPALVDW-----OSNNTFASCS 326
DB 289 WYNTWALSTDYALRTGAPEAPASVNPDLQSLQELKERALSRYNLVRGQGPRLVSGS 348
QY 327 TDMCIHVCKLQGD-RPIKTFQGHTEVNAIKWDPTGNILLASGDDMTLKIWSMKQDCNVH 385
DB 349 DDFTLFLNSPADKXPLRTMTGQALINQVLFSPDSRVASASFDKSLKWDGRTGKYL 408
QY 386 DLQQHNKEIYTIKWSPTGFGTNNPNANMLASASFDSTVRLWDVDRGICHTLTGHOEPV 445
```

```
Db 409 SURGHVAAYVQIAMSAD-----SRLVSGSDSTIKVWDVKAQKLANDLFGHADEV 459
QY 446 YSVAFSPDGRYLASGFDKCVHIW 469
DB 460 YAVDWSPDGQRVASGKDKCLRIV 483
RESULT 12
HUS7.MOUSE
ID HUS7.MOUSE STANDARD; PRT; 485 AA.
AC Q8VEJ4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WD-repeat protein HUSY-07.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Blakesley A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC018399; AAH18399.1; -
DR InterPro; IPR001632; Gpotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 7.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS0082; WD REPEATS 2; 7.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Nuclear protein; Repeat; WD repeat.
FT REPEAT 112 151
FT REPEAT 154 193
FT REPEAT 197 241
FT REPEAT 244 282
FT REPEAT 327 366
FT REPEAT 370 409
```

FT REPEAT 412 451 WD 7.
 FT REPEAT 454 484 WD 8.
 SQ SEQUENCE 485 AA: 53148 MN, 2D7F59D603AEC07B CRC64;
 Query Match 13.4%; Score 365.5; DB 1; Length 485;
 Best Local Similarity 26.6%; Pred. No. 3.2e-19;
 Matches 102; Conservative 48; Mismatches 145; Indels 89; Gaps 9;
 QY 165 LRCHSEVFICANPVSDLLASGSDSTARINWLSN-----STSGSTQLVLRHCIREGQ 220
 Db 110 LEHSEAVISVAFSPGKYLASGSDTTRFDLSTETPHFTCKGRRHWLISISPDGK 169
 QY 221 DVPS-----NKDVTSLDW-----NSEGLLATGSDVGFAR 250
 Db 170 KLASGCKNGQILLWDPSTGLQVTRTGLSHKMITGLSWEPLHNPCECRYVASSSDGYSR 229
 QY 251 IW-TKGNLSTLQGHKGFIFALKWKKGNFILSAGVDKTTIWDHATGEAKQPFPH-- 307
 Db 230 VMDTTAGRCERILTGHSTQVTCRLWGGDG-LIYSASQDRTIKVRAHDVLCRTLQGHCH 288
 QY 308 --SAPALVDW-----QSNNTFASCS 326
 Db 289 WNTWMLSTIDYALRTGAFEPAEATVNAQDLQGLSKELKERASSRYNLVRGQGPRLVSGS 348
 QY 327 TDCIHKVCKGQD-RIKTFQGHTEVNAIKWDPTGNILLASCSDDMTLKWMSKQDNCVH 385
 Db 349 DFTFLWSPAEKCKPLARMTGHQALINQVLFSPDSRIVASASFDKSLKMDGRTGKYLA 408
 QY 386 DLOQHNKEIYTIKWSPTGPTNNPNANLMLASFPDSTVRLMDVDRGICHTLTQHPEVP 445
 Db 409 SLRGHVAATQIWSAD-----SRLVSGSSDSTLKVWDYKAQLATDLPGHADVEV 459
 QY 446 YSAFSPDGRLYASGDPKCVHIW 469
 Db 460 YAVDWSPDGQVAVSGGKCLRIW 483
 RESULT 13
 LIS1 HUMAN STANDARD; PRT; 409 AA.
 ID LIS1 HUMAN STANDARD; PRT; 409 AA.
 AC P43034; Q8WZ88; Q8WZ89;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Platelet-activating factor acetylhydrolase IB alpha subunit (PAF
 DE acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha
 DE (PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1
 DE (PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1 OR PAPAH1B1
 OS Homo sapiens (Human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RN TISSUE=Brain, and Kidney;
 RX MEDLINE=93361119; PubMed=8355785;
 RA Reiner O., Carrozzo R., Shen Y., Wehnert M., Faustiniella F.,
 RA Dobyns W.B., Caskey C.T., Ledbetter D.H.,
 RT "Isolation of a Miller-Dieker lissencephaly gene containing G protein
 RT beta-subunit-like repeats.";
 RN Nature 364:717-721(1993).
 [2]
 RN SEQUENCE FROM N.A., AND VARIANT LIS1 ARG-148.
 RX MEDLINE=97217774; PubMed=9063735;
 RA Lo Nigro C., Chong S.S., Smith A.C.M., Dobyns W.B., Carrozzo R.,
 RA Ledbetter D.H.,
 RT "Point mutations and an intragenic deletion in LIS1, the lissencephaly
 RT causative gene in isolated lissencephaly sequence and Miller-Dieker
 RT syndrome.";
 RN Hum. Mol. Genet. 6:157-164 (1997).
 [3]
 RN SEQUENCE FROM N.A.
 RN TISSUE=Liver;
 RP SEQUENCE FROM N.A.
 RC

RA Zhao M.J., Xia S.L., Li T.P.;
 RT "High expression of the lissencephaly gene in hepatocarcinoma
 RT patients.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP Feng Z., Zhang B., Peng X., Yuan J., Qiang B.;
 RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [5]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Colon;
 RC Wambutt R., Heubner D., Mewes H.-W., Weil B., Amid C., Osanger A.,
 RA Pobo G., Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [6]
 RN SUBCELLULAR LOCATION, AND INTERACTION WITH DYNEIN AND DYNACTIN.
 RP MEDLINE=20512525; PubMed=11056532;
 RA Faulkner N.E., Dujardin D.L., Tai C.-Y., Vaughan K.T., O'Connell C.B.,
 RA Wang Y.-L., Vallee R.B.;
 RT "A role for the lissencephaly gene LIS1 in mitosis and cytoplasmic
 RT dynein function.";
 RL Nat. Cell Biol. 2:784-791 (2000).
 [7]
 RN INTERACTION WITH RSN; DYNEIN AND DYNACTIN.
 RP MEDLINE=21898370; PubMed=11889140;
 RA Tai C.-Y., Dujardin D.L., Faulkner N.E., Vallee R.B.;
 RT "Role of dynein, dynactin, and CLIP-170 interactions in LIS1
 RT kinetochore function.";
 RL J. Cell Biol. 156:959-968 (2002).
 [8]
 RN SUBCELLULAR LOCATION, AND INTERACTION WITH RSN.
 RP MEDLINE=21938534; PubMed=11940666;
 RA Coueille F.M., Caspi M., Cordelières F.P., Dompierre J.P.,
 RA Dujardin D.L., Koifman C., Martin P., Hoogenraad C.C., Akhmanova A.,
 RA Galjart N., De Mey J.R., Reiner O.;
 RT "LIS1, CLIP-170's key to the dynein/dynactin pathway.";
 RL Mol. Cell. Biol. 22:3089-3102 (2002).
 [9]
 RN VARIANT SBH PRO-168.
 RP MEDLINE=99371784; PubMed=10441340;
 RA Pilz D.T., Kuc J., Matsumoto N., Bodurtha J., Bernadi B.,
 RA Tassinari C.A., Dobyns W.B., Ledbetter D.H.;
 RT "Subcortical band heterotopia in rare affected males can be caused by
 RT missense mutations in DCX (XLIS) or LIS1.";
 RN Hum. Mol. Genet. 8:1757-1760 (1999).
 [10]
 RN VARIANTS LIS1 SER-30; SER-161 AND HIS-316.
 RP MEDLINE=21394321; PubMed=11502906;
 RA Leventer R.J., Cardoso C., Ledbetter D.H., Dobyns W.B.;
 RT "LIS1 missense mutations cause milder lissencephaly phenotypes
 RT including a child with normal IQ.";
 RL Neurology 57:416-422 (2001).
 CC -!- FUNCTION: Probably involved in nuclear migration during cell
 CC division. Participates in the process of neuronal cell
 CC differentiation or brain development, possibly by through its role
 CC in nuclear migration of differentiating neurons. May participate
 CC in targeting cytoplasmic dynein to the microtubule plus ends,
 CC thereby playing an essential role in dynein-mediated microtubule
 CC sliding. Non-catalytic subunit of an acetylhydrolase complex, a
 CC complex that inactivates platelet-activating factor (PAF) by
 CC removing the acetyl group at the SN-2 position.
 CC -!- SUBUNIT: Cytosolic PAF-AH IB is formed of three subunits of 45 kDa
 CC (alpha), 30 kDa (beta) and 29 kDa (gamma). The catalytic activity
 CC of the enzyme resides in the beta and gamma subunits, whereas the
 CC alpha subunit has regulatory activity. Trimer formation is not
 CC essential for the catalytic activity. Interacts with dynein and
 CC dynactin. Interacts with RSN.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; localizes to the plus end of
 CC microtubules.
 CC -!- TISSUE SPECIFICITY: Fairly ubiquitous expression in both the
 CC frontal and occipital areas of the brain.
 CC -!- DISEASE: Defects in PAPAH1B1 are the cause of classical
 CC lissencephaly-1 (LIS1) [MIM:607432]. LIS1 is a brain malformation

caused by abnormal neuronal migration at 9 to 13 weeks' gestation. Lissencephaly means 'smooth brain', a brain without convolutions or gyri, and consists in an absence (agyria) or a decrease (pachygyria) of surface convolutions associated with a disorganization of the clear neuronal lamination of normal six-layered cortex. It is characterized by an abnormally thick and poorly organized cortex with 4 primitive layers, diffuse neuronal heterotopia, enlarged and dysmorphic ventricles, and often hypoplasia of the corpus callosum.

-1- DISEASE: Defects in PAFAH1B1 are the cause of subcortical band heterotopia (SBH) [MIM:607432]. SBH is at the less severe end of the lissencephaly spectrum of malformations characterized by bilateral and symmetric ribbons of gray matter found in the central white matter between the cortex and the ventricular surface.

-1- DISEASE: Defects in PAFAH1B1 are a cause of Miller-Dieker lissencephaly syndrome (MDLS) [MIM:247200]. MDLS is a form of lissencephaly associated with facial abnormalities.

-1- SIMILARITY: Contains 1 LISH domain.

-1- SIMILARITY: Contains 7 WD repeats.

-1- CAUTION: Ref.1 (AA02882) sequence differs from that shown due to a chimeric cDNA.

-1- CAUTION: Ref.5 sequence differs from that shown due to what seems to be the presence of intronic sequence in the cDNA.

-1- DATABASE: NAME=HotMolbase; NOTE=LISI entry;

WWW="http://bioinformatics.weizmann.ac.il/hotmolbase/entries/lisi.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L13385; AA02880.1; -
 EMBL; L13386; AA02881.1; -
 EMBL; L13387; AA02882.1; ALT_SEQ.
 EMBL; U72342; AAC51111.1; -
 EMBL; U72334; AAC51111.1; JOINED.
 EMBL; U72335; AAC51111.1; JOINED.
 EMBL; U72336; AAC51111.1; JOINED.
 EMBL; U72337; AAC51111.1; JOINED.
 EMBL; U72338; AAC51111.1; JOINED.
 EMBL; U72339; AAC51111.1; JOINED.
 EMBL; U72340; AAC51111.1; JOINED.
 EMBL; U72341; AAC51111.1; JOINED.
 EMBL; AF208837; AAL34972.1; -
 EMBL; AF208838; AAL34973.1; -
 EMBL; AF400434; AAK92483.1; -
 EMBL; BX538346; CAD98141.1; -
 Genew; HGNC:8574; PAFAH1B1.
 MIM; 601545; -
 MIM; 607432; -
 MIM; 247200; -

GO; GO:0006928; P:cell motility; TAS.

GO; GO:0006629; P:lipid metabolism; TAS.

GO; GO:0007399; P:neurogenesis; TAS.

GO; GO:0007165; P:signal transduction; TAS.

InterPro; IPR006594; Lish.

InterPro; IPR001680; WD40.

Pfam; PF00400; WD40; 7.

PRINTS; PR00320; GPROTEINBRPT.

ProDom; PD000018; WD40; 5.

SMART; SM00667; Lish; 1.

SMART; SM00320; WD40; 7.

PROSITE; PS00896; LISH; 1.

PROSITE; PS00678; WD_REPEATS_1; 4.

PROSITE; PS00682; WD_REPEATS_2; 7.

PROSITE; PS00294; WD_REPEATS_REGION; 1.

Cell division; Mitosis; Neurogenesis; Cytoskeleton; Microtubule;

Repeat; Coiled coil; WD repeat; Disease mutation; Lissencephaly.

INIT_MET 0 BY SIMILARITY.

FT	DOMAIN	6	38	LISH.
FT	DOMAIN	50	81	COILED COIL (POTENTIAL).
FT	REPEAT	105	135	WD 1.
FT	REPEAT	147	177	WD 2.
FT	REPEAT	189	219	WD 3.
FT	REPEAT	231	261	WD 4.
FT	REPEAT	273	303	WD 5.
FT	REPEAT	315	345	WD 6.
FT	REPEAT	357	387	WD 7.
FT	VARIANT	30	30	F -> S (in LISI).
FT	VARIANT	148	148	/FTid=VAR_015398.
FT	VARIANT	161	161	H -> R (in LISI).
FT	VARIANT	168	168	/FTid=VAR_007724.
FT	VARIANT	168	168	G -> S (in LISI).
FT	VARIANT	168	168	/FTid=VAR_015399.
FT	VARIANT	316	316	S -> P (in SBH).
FT	VARIANT	316	316	/FTid=VAR_010203.
FT	VARIANT	316	316	D -> H (in LISI).
FT	VARIANT	316	316	/FTid=VAR_015400.
FT	CONFLICT	20	20	S -> P (IN REF. 3).

Query Match 12.9%; Score 351.5; DB 1; Length 409;
 Best Local Similarity 22.2%; Pred. No. 2.6e-18;
 Matches 109; Conservative 77; Mismatches 165; Indels 141; Gaps 15;

QY	6	DEVNLFVRYLQESGFHSHTFTGKSHISOSNINGALVPPAALISITQKGLQVVEA	65
QY	6	DEVNLFVRYLQESGFHSHTFTGKSHISOSNINGALVPPAALISITQKGLQVVEA	65
DB	8	DELNRAIADYLRNGYEAE	37
QY	66	INEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAVRDQLAQOQAAAAAASQGS	125
DB	38	VNEE---LDKK---YAGLEKKWTSVRLQKKNVELESKNEKEEFTSGGFLQKRD	91
QY	126	NGENTANGEANGAHTIANHHTDMMEVDGVEIPPKNKAVVLEHSEVPICAWNPVSD	185
DB	92	EW-----IPRPPEK-YALSCHRSPTVTRVIFHEVFSVMV	123
QY	186	SGSGDSTARINLNS---ENSTSGTQLVLEHCTREGQDVPSNKDVTSLDWNSEG	241
DB	124	SASEDAIKWDYETGDFERTLKHTDSV-----QDI-----SFDHSGKLLA	165
QY	242	TGSYDGFARIWKDG-NLASTLQGHKGFIFALKWNKGNFTLSAGVDTTIINDAHT	300
DB	166	SCSADMTIKLWDFQFECIRIMHGHENSVVAIMPNGDHIVASRDKTIKWEVQIG	223
QY	301	KQPFPSAPALVDWQSNFTFSCSDMCIHVCKLQCDRPIKTFQGHTEVNAIKWDPT	360
DB	224	-----YC-----VKTFTHREWVRVRENQD	244
QY	361	GNLLASGDDMTLKIWSMKQNCVHDLQCHNKETVYTIKWSP-----TGPCT-NNP	409
DB	245	GTLLASCSNDQTVRVVVVATKECAELREHVVVEICISWAPESYSISATGSETKSG	304
QY	410	NANMLASAFDSTVRLWDVDRGICHTLTKHQPVEVYVAFSPDGRYLASGDFKCVH	469
DB	305	KPGFLLSGSRDKTIKWMDVSTGCMCLMTLVGHNDVVRGVLVFGSGKFLSCADDK	364
QY	470	N-TQVCLHYLN	479
DB	365	DYKNGRCNKTLN	376

RESULT 14

LISI_BOVIN

ID LISI_BOVIN STANDARD; PRT; 409 AA.

AC P43033;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Platelet-activating factor acetylhydrolase IB alpha subunit (PAF

DE acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa subunit) (PAF-AH alpha)

DE (PAF-AH alpha) (Lissencephaly-1 protein) (LISI-1).

GN PAFAH1B1 OR PAF-AH OR LISI.

8 DELNRAIADYLRNGYE-----AAYSVFKK-----EAEID 37

66 INEDGTFLDGRPIESLSDAVMPDVVQTRQAYRDKLAQQQAAAAAASQGSAAK 125

38 MNES--LDDK--YAGLLEKWTSVIRLQKKWELSKNEAKEFTSGGPLGQKRDPK 91

126 NGENTANGENGARTIANNHDTMWEDVDGVEIPPKAVVLRGHESVFFICAWNPNVSDLLA 185

92 EW-----IPRPPEK-YALSGHRSPVTRVIFHPVFSVMV 123

186 SSGSGDSTARINWLS-----ENSTSGSTQLVLRHCIRREGQGDVPSNKDVTSLDWNSECTLLA 241

124 SASEDATIKWYDETGFESTLKGHTD-----SVEDISFDHSGKLLA 165

242 TGSVDGPARIWTXDG-NLASTLQGHKGPIFALKWNKKGFIISAGYDKTTIILWDAHTGEA 300

166 SCSADMIYKLDWFGQFECIRTMGHGDHNVSSVALMPGDHIVSASRDKTIKMWVEVQTG-- 223

301 KQQPFHSHAPALVDWQSNNTFASCSTDCMCIHVCKLQDQPIKTFQGHTEWNAIKWDPT 360

224 -----YC-----VKTFGHRWVWVRPNQD 244

361 GNLLASCSDDMTLKIWMKQDNCVHDLQOHNKEIYTIKMS-----TGPGT-NNP 409

245 GTLTASCNDQTVRVVVATKECKAELREHEHVVEICISWAPESSYSISBATSETKKSG 304

410 NANMLASASPDSTVRWDVDRGICHTLTKHQEPVYSVAFSPDGYVLASGSPDKCVHIW 469

305 KPGPFLLSGSRDKTIKWDVSTGMCLMTLVGHNDWVRGVLFFHSGGRFILLSCADDKTLRW 364

470 N--TQVCLHYLN 479

365 DYKNKRCMKTLN 376

RESULT 15

CIAL HUMAN

ID CIAL HUMAN STANDARD; PRT; 339 AA.

AC O76071;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE WD-repeat containing protein Ciao 1.

GN CIAL01.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.

RP MEDLINE=59825157; PubMed=9556563;

RX Johnstone R.W., Wang J., Tommerup N., Vissing H., Roberts T., Shi Y.,

RT "Ciao 1 is a novel WD40 domain protein that interacts with the tumor

RL suppressor protein WT1";

RT J. Biol. Chem. 273:10880-10887(1998).

RN [2]

RN SEQUENCE FROM N.A.

RP Adams M.D., Loftus B.J., Zhou L., Crosby M., Fuhrmann J., Mason T.M.,

RP Brandon R., Kim U.J., Kerlavage A.R., Venter J.C.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RP TISSUE=Blood, and Skin;

RC MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datta-Chencho L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Skowronski M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villan J.D., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4].
 RP CHARACTERIZATION.
 RX MEDLINE=93367400; PubMed=10438340;
 RA Johnstone R.W., Tommerup N., Hansen C., Vissing H., Shi Y.;
 RT "Structural organization, tissue expression, and chromosomal
 RT localization of Cioa 1, a functional modulator of the Wilms' tumor
 RT suppressor, Wt1.";
 RL Immunogenetics 49:900-905 (1999).
 CC -!- FUNCTION: Seems to specifically modulate the transactivation
 CC activity of Wt1.
 CC -!- SUBUNIT: Interacts with Wt1.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- MISCELLANEOUS: 'Ciao' means 'bridge' in Chinese.
 CC -!- SIMILARITY: Contains 7 WD repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; U63810; AAC234948.1; --
 CC EMBL; AC004020; AAC23493.1; --
 CC EMBL; BC001395; AAH01395.1; --
 CC EMBL; BC032812; AAH32812.1; --
 CC MIM; 604333; --
 CC GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 CC GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 7.
 CC PRINTS; PR00320; GPROTEINBRPT.
 CC ProDom; PD000018; WD40; 2.
 CC SMART; SM00320; WD40; 7.
 CC PROSITE; PS00678; WD_REPEATS_1; 1.
 CC PROSITE; PSS0082; WD_REPEATS_2; 6.
 CC PROSITE; PSS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Nuclear protein.
 FT REPEAT 14 53 WD 1.
 FT REPEAT 59 98 WD 2.
 FT REPEAT 103 142 WD 3.
 FT REPEAT 148 187 WD 4.
 FT REPEAT 192 231 WD 5.
 FT REPEAT 250 289 WD 6.
 FT REPEAT 301 338 WD 7.
 SQ SEQUENCE 339 AA; 37840 MW; 63A8D8257A204FC8 CRC64;
 Query Match 12.9%; Score 350; DB 1; Length 339;
 Best Local Similarity 31.9%; Pred. No. 2.6e-18;
 Matches 86; Conservative 79; Indels 64; Gaps 9;
 QY 222 VPENKDVTS--LWNSEGTLTAGSDVGFIARIWTKDGN---LASTLQGHKGPIFALKWN 275
 DB 11 VPAHPDSRCFLAWNFAGTLTASCGGDRRIIRIWGTEDGSWICKSVLSEGHQRTVRKVAWS 70
 QY 276 KKGNFILSAGVDKTTIWDHAHTGEAKQPFPHSAPALDVDWQSNNTFASCTDMCHVCK 335
 DB 71 PCGNYLASFDATTCTI-----WKKNQDDFCVTT-----99
 QY 336 LGQDRPIKTFQGHNEVNAIKWDP TGNLLASCSDDDMTLKIWSMKQDN---CVHDLQHNK 392

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 2718

Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLWVPLVALIELLVK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2468	90.8	514	11	Q8BHJ5	Q8BHJ5 mus musculus
2	2465	90.7	514	11	Q8CBG4	Q8CBG4 mus musculus
3	2464	90.7	514	11	Q9EQD4	Q9EQD4 mus musculus
4	2421.5	89.1	519	13	Q7SZM9	Q7SZM9 xenopus lae
5	2273.5	83.6	527	11	Q8BYQ4	Q8BYQ4 mus musculus
6	2269.5	83.5	527	11	Q8BMM0	Q8BMM0 mus musculus
7	2229	82.0	577	4	Q8GUV2	Q8GUV2 homo sapien
8	1761.5	64.8	412	11	Q8COA1	Q8COA1 mus musculus
9	1715.5	63.1	700	5	Q9XZK1	Q9XZK1 drosophila
10	1251	46.0	613	10	Q9FN19	Q9FN19 arabidopsis
11	923.5	34.0	524	5	Q9SRJ9	Q9SRJ9 drosophila
12	882	32.5	201	11	Q8VEG3	Q8VEG3 mus musculus
13	444.5	16.4	1227	16	Q8ZOR1	Q8ZOR1 anabaena sp
14	442	16.3	1356	3	Q8X1P3	Q8X1P3 podospora a
15	442	16.3	1356	3	Q8X1P5	Q8X1P5 podospora a
16	433.5	15.9	1356	3	Q8X1P4	Q8X1P4 podospora a

17	429.5	15.8	1189	16	Q8YLO9	Q8YLO9 anabaena sp
18	426	15.7	1708	16	Q8YZI2	Q8YZI2 anabaena sp
19	423.5	15.6	1376	3	Q8X1P2	Q8X1P2 podospora a
20	408	15.0	1711	16	Q8Z019	Q8Z019 anabaena sp
21	404.5	14.9	934	16	Q8YZ23	Q8YZ23 anabaena sp
22	395	14.5	1747	16	Q8Z020	Q8Z020 anabaena sp
23	382.5	14.1	564	3	Q74845	Q74845 schizosacch
24	381	14.0	1551	16	Q8YMU3	Q8YMU3 anabaena sp
25	376	13.8	415	4	Q8N136	Q8N136 homo sapien
26	370.5	13.6	411	5	Q96698	Q96698 drosophila
27	369	13.6	676	16	Q8YSG6	Q8YSG6 anabaena sp
28	368.5	13.6	304	11	Q8D4T2	Q8D4T2 mus musculu
29	368.5	13.6	304	16	Q8Z0F4	Q8Z0F4 anabaena sp
30	368	13.5	415	4	Q8N776	Q8N776 homo sapien
31	367	13.5	410	13	Q803D2	Q803D2 brachydanio
32	366	13.5	502	3	Q74855	Q74855 schizosacch
33	365	13.4	1189	16	Q8YTD1	Q8YTD1 anabaena sp
34	360.5	13.3	410	13	Q902L4	Q902L4 xenopus lae
35	357.5	13.2	1241	2	Q9XB08	Q9XB08 amycolatops
36	356	13.1	352	10	Q80990	Q80990 arabidopsis
37	354	13.0	265	16	Q8YZ16	Q8YZ16 anabaena sp
38	353.5	13.0	476	13	Q93531	Q93531 xenopus lae
39	353.5	13.0	476	13	Q7ZXK9	Q7ZXK9 xenopus lae
40	352	13.0	339	11	Q9DCZ7	Q9DCZ7 mus musculu
41	352	13.0	339	11	Q99KN2	Q99KN2 mus musculu
42	351.5	12.9	1430	16	Q98HK1	Q98HK1 rhizobium l
43	349.5	12.9	410	6	Q9GL51	Q9GL51 sus scrofa
44	349.5	12.9	410	13	Q9PTR5	Q9PTR5 gallus gall
45	344.5	12.7	559	16	Q8YSC0	Q8YSC0 anabaena sp

ALIGNMENTS

RESULT 1

Q8BHJ5 PRELIMINARY; PRT; 514 AA.
AC Q8BHJ5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE IRAI protein.
GN IRAI OR 8030499H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK029595; BAC26526.1; -
DR EMBL; AK033347; BAC28241.1; -
DR PIR; PT0651; PT0651.
DR MGD; MGI:2441730; Ira1.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOT; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS0082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 90.8%; Score 2468; DB 11; Length 514;
 Best Local Similarity 98.5%; Pred. No. 5e-167;
 Matches 465; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOBSGFSGHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOBSGFSGHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAATNQ 120

QY 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180

QY 181 SLLASGSGDSTARIMWLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLASGSGDSTARIMWLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

QY 241 ATGSYDGFARITWKDGNLASTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLASTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300

QY 301 KQPFPHSAPALVDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360
 DB 301 KQPFPHSAPALVDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKETIYIKWSPGPTGPTNNPNNMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKETIYIKWSPGPTGPTNNPNNMLASASF 420

QY 421 DSTVRLMDVDRGICHTLTTHKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472
 DB 421 DSTVRLMDVDRGICHTLTTHKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472

RESULT 2

Q8CBG4 PRELIMINARY; PRT; 514 AA.

ID Q8CBG4
 AC Q8CBG4
 DT 01-VAR-2003 (Tremblrel. 23, Created)
 DT 01-VAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RC MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK036064; BAC29294.1; -
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD REPEATS 1; 4.
 DR PROSITE; PS50082; WD REPEATS 2; 6.
 DR PROSITE; PS50294; WD REPEATS_REGION; 1.

SQ SEQUENCE 514 AA; 55689 MW; 13BED3753A725029 CRC64;

Query Match 90.7%; Score 2465; DB 11; Length 514;
 Best Local Similarity 98.3%; Pred. No. 8.2e-167;
 Matches 464; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOBSGFSGHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOBSGFSGHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAATNQ 120

QY 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAXNGENTANGEANGAHTIANNHTDMVEVDGVEIPNKAVVLRGHESEVFICAWNPV 180

QY 181 SLLASGSGDSTARIMWLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLASGSGDSTARIMWLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

QY 241 ATGSYDGFARITWKDGNLASTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSYDGFARITWKDGNLASTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300

QY 301 KQPFPHSAPALVDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360
 DB 301 KQPFPHSAPALVDVQSNNTFASCSTDMCHVCKLGQDRPIKTFQHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKETIYIKWSPGPTGPTNNPNNMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKETIYIKWSPGPTGPTNNPNNMLASASF 420

QY 421 DSTVRLMDVDRGICHTLTTHKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472
 DB 421 DSTVRLMDVDRGICHTLTTHKQEPVSVAFSPDGRYLASGDFKCVHIWNTQ 472

RESULT 3

Q8EQD4 PRELIMINARY; PRT; 514 AA.

ID Q8EQD4
 AC Q8EQD4
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE IRAL.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Zhang X.; Dormady S.; Basch R.;
 RT "Identification of four human cDNAs that are differentially expressed
 RT by early hematopoietic progenitors."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268195; AAG4738.1; -
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; Lish; 1.
 DR PROSITE; PS00678; WD REPEATS 1; 4.
 DR PROSITE; PS50082; WD REPEATS 2; 6.
 DR PROSITE; PS50294; WD REPEATS_REGION; 1.
 KW Repeat; WD repeat.

SQ SEQUENCE 514 AA; 55689 MW; 6A72CE68A40C141F CRC64;

Query Match
Best Local Similarity 90.7%; Score 2464; DB 11; Length 514;
Matches 464; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISDEVNVLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAASQ 120
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAAATNQ 120

QY 121 QGSAKNGENTANGENGCAHTIANNHTDMEVDGVEIPPNKAVILRGHSEVFICAWNPV 180
DB 121 QGSAKNGENTANGENGCAHTIANNHTDMEVDGVEIPPNKAVILRGHSEVFICAWNPV 180

QY 181 SLLVSGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKKDVTSLDWSSEGTLL 240
DB 181 SLLVSGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKKDVTSLDWSSEGTLL 240

QY 241 ATGSDYGFARITWKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300
DB 241 ATGSDYGFARITWKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300

QY 301 KQGFPPHSPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFOGHTNEVNAIKWDPT 360
DB 301 KQGFPPHSPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFOGHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASASF 420
DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASASF 420

QY 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472

RESULT 4
Q7SZM9 PRELIMINARY; PRT; 519 AA.

AC Q7SZM9
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor {alpha} with Promyelocytic Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in Vivo.";
RL J. Biol. Chem. 278:30788-30795(2003).
DR EMBL; AY225088; AAP20646.1; -;
KW Receptor.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match
Best Local Similarity 96.0%; Score 2421.5; DB 13; Length 519;
Matches 458; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISDEVNVLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAA-- 118
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQAQAAAAA 120

QY 119 ---SQGSAKNGENTANGENGCAHTIANNHTDMEVDGVEIPPNKAVILRGHSEVFIC 175
DB 121 TPNQOPPAKNGENTANGENGCAHTIANNHTDMEVDGVEIPPNKAVILRGHSEVFIC 180

QY 176 AWPVSDLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKKDVTSLDWS 235
DB 181 AWPVSDLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKKDVTSLDWS 240

QY 236 EGTLLATGSDYGFARITWKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDA 295
DB 241 EGTLLATGSDYGFARITWKDGNLSTLQGHKGPFPALKWKNKGNFILSAGVDKTTIIWDA 300

QY 296 HTGEAKQGFPPHSPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFOGHTNEVNAI 355
DB 301 HTGEAKQGFPPHSPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFOGHTNEVNAI 360

QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANML 415
DB 361 KWDPTGNLLASCSDDMTLKIWSMKHDTCTVHDLQAHNKEIYTIKWSPTGPGTNNPNANML 420

QY 416 ASASFDSVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 472
DB 421 ASASFDSVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQ 477

RESULT 5
Q8BYQ4 PRELIMINARY; PRT; 527 AA.

AC Q8BYQ4
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Transducin (Tblix protein).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Joqueillano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

RT and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN (3).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RA Strausberg R.; 2003) to the EMBL/GenBank/DBJ databases.
RL Submitted (JAN-2003) BAC30092.1; -.
DR EMBL; AK038674; BAC30092.1; -.
DR EMBL; BC043105; AAH43105.1; -.
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;

Query Match 83.6%; Score 2273.5; DB 11; Length 527;
Best Local Similarity 87.4%; Pred. No. 3.4e-153;
Matches 425; Conservative 26; Mismatches 20; Indels 15; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISDEVNLFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAFREKLTQQQANAAAAAAT 120
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAFREKLTQQQANAAAAAAT 120
QY 116 -----AAASQGSANGENTANGENGAAHTIANNHTDMMEVDGVEIPPNAVVL 166
DB 121 ATSTAATPAAQAQNPKNKEATVNGEENGAAHAI-NNHSKPWEIDGVEIPPNAVVL 179
QY 167 GHESEVFICAWNPSVSDLLASGSDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFICAWNPSVSDLLASGSDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNK 239
QY 227 DVTSLDWNSDGLTATGSDYDGFARITWTKDGNLSTLGOHKGPIFALKWNKKNFILSAGV 286
DB 240 DVTSLDWNSDGLTATGSDYDGFARITWTKDGNLSTLGOHKGPIFALKWNKKNFILSAGV 299
QY 287 DKTTIWDHAHTGEAKQOPPHSAPALDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDHAHTGEAKQOPPHSAPALDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359
QY 347 GHTNEVNAIKWDPPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 406
DB 360 GHTNEVNAIKWDPPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 419
QY 407 NNPNANMLASASFDSTVRLWDVDRGICITHTLTKHQPVPVSVAFSPDGRYLASGDFKCV 466
DB 420 SNPNANMLASASFDSTVRLWDVDRGICITHTLTKHQPVPVSVAFSPDGRYLASGDFKCV 479
QY 467 HIWNTQ 472
DB 480 HIWNTQ 485

RESULT 6
Q8BMMO PRELIMINARY; PRT; 527 AA.
ID Q8BMMO
AC Q8BMMO;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Transducin.

GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs." ;
DR EMBL; AK030547; BAC27015.1; -.
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
SQ SEQUENCE 527 AA; 56832 MW; 0466F764ABA25CB0 CRC64;

Query Match 83.5%; Score 2269.5; DB 11; Length 527;
Best Local Similarity 87.2%; Pred. No. 6.6e-153;
Matches 424; Conservative 26; Mismatches 21; Indels 15; Gaps 2;

QY 1 MSISDEVNLFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
DB 1 MSISDEVNLFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
QY 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAFREKLTQQQANAAAAAAT 120
DB 61 EAEVISEDGTLFDGRPIESLSLIDAVMPDVVQTRQAFREKLTQQQANAAAAAAT 120
QY 116 -----AAASQGSANGENTANGENGAAHTIANNHTDMMEVDGVEIPPNAVVL 166
DB 121 ATSTAATPAAQAQNPKNKEATVNGEENGAAHAI-NNHSKPWEIDGVEIPPNAVVL 179
QY 167 GHESEVFICAWNPSVSDLLASGSDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFICAWNPSVSDLLASGSDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNK 239
QY 227 DVTSLDWNSDGLTATGSDYDGFARITWTKDGNLSTLGOHKGPIFALKWNKKNFILSAGV 286
DB 240 DVTSLDWNSDGLTATGSDYDGFARITWTKDGNLSTLGOHKGPIFALKWNKKNFILSAGV 299
QY 287 DKTTIWDHAHTGEAKQOPPHSAPALDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDHAHTGEAKQOPPHSAPALDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359
QY 347 GHTNEVNAIKWDPPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 406
DB 360 GHTNEVNAIKWDPPTGNLLASCDSDMTLKIWSMKQDNCVHDLQOHKEIYTIKNSPTGPT 419
QY 407 NNPNANMLASASFDSTVRLWDVDRGICITHTLTKHQPVPVSVAFSPDGRYLASGDFKCV 466
DB 420 SNPNANMLASASFDSTVRLWDVDRGICITHTLTKHQPVPVSVAFSPDGRYLASGDFKCV 479
QY 467 HIWNTQ 472
DB 480 HIWNTQ 485

RESULT 7
Q86UY2

```

```

ID AC Q86UY2 PRELIMINARY; PRT; 577 AA.
AC Q86UY2;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE TBL1X protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshlyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Shevchenko Y., Bouffard G.G.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052304; AAH52304.1;
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS50896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 6.
DR PROSITE; PS50294; WD REPEATS 2; 6.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

Query Match 82.0%; Score 2229; DB 4; Length 577;
Best Local Similarity 85.4%; Pred. No. 5.6e-150;
Matches 414; Conservative 32; Mismatches 25; Indels 14; Gaps 2;

QY 1 MSISDEVNFLVYRLQSGFGSHSAFTGKISHISQSNINGALVPPAALISITQKGLQYV 60
Db 52 MSITSDVNFVYRLQSGFGSHSAFTGISHISQSNINGTLVPPAALISITQKGLQYV 111
QY 61 EAEVSIINDEGTLFDGRPIESLSLIDAVMPDVVQTSQAYRDKLAQQQAAAAAATAA 117
Db 112 EAEISINDEGTVFGRPIESLSLIDAVMPDVVQTSQAYRDKLAQQQAAAAAATAA 171
QY 118 -----ASQGSAGKNGENTANGAHTIANNHTDMWVDGVDVEIPNKAVLVR 167
Db 172 ATAATTAGVSHQNPKNREATVNGEENRAHSV--NNHAKPWEIDGVEIIPSKATVLR 230
QY 168 HESEVFICAMPVPDILLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQVPSNKD 227
Db 231 HESEVFICAMPVPDILLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQVPSNKD 290

```

```

QY 228 VTSLDWNSEGTLLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIISAGVD 287
Db 291 VTSLDWNSEGTLLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIISAGVD 350
QY 288 KTTIIWDAHTGEAKQOFPFHSAPALVDVQSNNTTASCSTDMCIHVCKLGQDRPKTFQG 347
Db 351 KTTIIWDAHTGEAKQOFPFHSAPALVDVQSNNTTASCSTDMCIHVCKLGQDRPKTFQG 410
QY 348 HTNEVNAIKWDPDPTGNLLASCSDDMTLKWSMKQDNCVHDLOQHNKEIYTIKNSPTGPGTN 407
Db 411 HTNEVNAIKWDPDPTGNLLASCSDDMTLKWSMKQDNCVHDLOQHNKEIYTIKNSPTGPGTN 470
QY 408 NFNAMLMLASASFDSTVRLMDVDRGICHTLTKHQPVPVYVAFSPDGRYLASGSPDKCVH 467
Db 471 NFNAMLMLASASFDSTVRLMDVDRGICHTLTKHQPVPVYVAFSPDGRYLASGSPDKCVH 530
QY 468 IWNTO 472
Db 531 IWNTO 535

RESULT 8
O8COA1 PRELIMINARY; PRT; 412 AA.
ID O8COA1;
AC O8COA1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transducin (Fragment).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK031937; BAC27612.1; -.
DR FIR; FT0651; FT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS50082; WD REPEATS 2; 6.
DR PROSITE; PS50294; WD REPEATS REGION; 1.
FT NON TER 1
SQ SEQUENCE 412 AA; 44356 MW; EB78910E6D9E5237 CRC64;

Query Match 64.8%; Score 1761.5; DB 11; Length 412;
Best Local Similarity 87.0%; Pred. No. 5.8e-117;
Matches 321; Conservative 20; Mismatches 27; Indels 1; Gaps 1;

QY 104 AQOQAAAAAASQGSAGKNGENTANGAHTIANNHTDMWVDGVDVEIPNKAV 163
Db 3 AATATTAATTAAQAQNPKNREATVNGEENRAHSV--NNHAKPWEIDGVEIIPSKAT 61
QY 164 VLRGHESEVFICAMPVPDILLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQVDP 223
Db 62 VLRGHESEVFICAMPVPDILLASGSDSTARIWNLSNSTSGTQLVLRHCIREGGQVDP 121
QY 224 SNKDVTSLDWNSGTLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIIS 283
Db 122 SNKDVTSLDWNSGTLATGSDYGFARIWTKDGNLASTLGOHKGPFPALKKWKNGNFIIS 181

```

QY 284 AGVDTKTIWDAHTGEAKQPPHSPAPALDVWDQSNNTFASCSTDMCIHVCKLQDORPIK 343
 Db 182 AGVDTKTIWDAHTGEAKQPPHSPAPALDVWDQSNNTFASCSTDMCIHVCKLQDORPIK 241
 QY 344 TFGHTNEVNAIKWPBTGNLLASCDDMTLTKWSKQDNCVHDLQKHNIKIYIKKSPG 403
 Db 242 TFGHTNEVNAIKWPBTGNLLASCDDMTLTKWSKQDNCVHDLQKHNIKIYIKKSPG 301
 QY 404 PGTNNPNAIMLASASFTSTVRLWVDVRCIGICHTLTKHQBPPVYVAFSPDGRVLAGSFD 463
 Db 302 PATSPNSNIMLASASFTSTVRLWVDVRCIGICHTLTKHQBPPVYVAFSPDGRVLAGSFD 361
 QY 464 KCVHIWNTQ 472
 Db 362 KCVHIWNTQ 370

RESULT 9
 Q9XZK1 PRELIMINARY; PRT: 700 AA.
 AC Q9XZK1; (TREMELrel. 12, Created)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE EBI protein.
 GN EBI OR CG4063.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazeg R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Fabros B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).

QY 1 MSISDENVLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIQKGLV 60
 Db 1 MSFSDENVLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIQKGLV 60
 QY 61 EAEVSTINEDGTLFGRPIESLSLIDAWPDV-----VOTR----- 96
 Db 61 EAEVSTINEDGTLFGRPIESLSLIDAWPDV-----VOTR----- 96
 QY 61 EAEVSTINEDGTLFGRPIESLSLIDAWPDV-----VOTR----- 96
 Db 61 EAEVSTINEDGTLFGRPIESLSLIDAWPDV-----VOTR----- 96
 QY 97 -----QAVRDKLA----- 130
 Db 97 -----QAVRDKLA----- 130
 QY 119 NAKPEIKIEPTGTVAGSAGNGKIAGTGTGTSPTDQSAEVDSSGNANAGTVAGNG 178
 Db 119 NAKPEIKIEPTGTVAGSAGNGKIAGTGTGTSPTDQSAEVDSSGNANAGTVAGNG 178
 QY 131 ANGE----- 135
 Db 131 ANGE----- 135
 QY 179 AGNGASTGSGNSTSTPAGGLAAGPAGKQKQSGNSAGSSSGNAGNANATSDAASS 238
 Db 179 AGNGASTGSGNSTSTPAGGLAAGPAGKQKQSGNSAGSSSGNAGNANATSDAASS 238
 QY 136 -----NG-----AHTIANNHTM----- 148
 Db 136 -----NG-----AHTIANNHTM----- 148
 QY 239 TSTNGNSTSSVEQPTSGITPAGGTGTVSTNPDAAAGGASTATGKAPSGAVTIRVGAQ 298
 Db 239 TSTNGNSTSSVEQPTSGITPAGGTGTVSTNPDAAAGGASTATGKAPSGAVTIRVGAQ 298
 QY 149 -----MEVDGVEIPPKAVVLRGHESEV 172
 Db 149 -----MEVDGVEIPPKAVVLRGHESEV 172
 QY 299 GNNVQSSSSNAQSAFSGTISSTSGAGTTPAALVPMIDENIEIPESKARVLRGHESEV 358
 Db 299 GNNVQSSSSNAQSAFSGTISSTSGAGTTPAALVPMIDENIEIPESKARVLRGHESEV 358
 QY 173 FICAWNPSVLLASGSGDSTARVNLSENSTSGTQLVLRHCIRREGGQDVPNSKDVTSID 232
 Db 173 FICAWNPSVLLASGSGDSTARVNLSENSTSGTQLVLRHCIRREGGQDVPNSKDVTSID 232
 QY 359 FICAWNPSVLLASGSGDSTARVNLSENSTSGTQLVLRHCIRREGGQDVPNSKDVTSID 417
 Db 359 FICAWNPSVLLASGSGDSTARVNLSENSTSGTQLVLRHCIRREGGQDVPNSKDVTSID 417
 QY 233 WNSEGTLLATGSDVGPARTWTKDGNLSTLQHKGPFPALKKNGKNTLSAGVDTII 292
 Db 233 WNSEGTLLATGSDVGPARTWTKDGNLSTLQHKGPFPALKKNGKNTLSAGVDTII 292

Query Match 63.1%; Score 1715.5; DB 5; Length 700;
 Best Local Similarity 52.9%; Pred. No. 2.3e-113;
 Matches 349; Conservative 49; Mismatches 71; Indels 191; Gaps 8;

```

Db 418 WNCDSGLATGSDGVARIWKTGRLASTLGQKGFIFALKWKNCGNYILSAGVDKTTII 477
QY 293 WDAHTGEAKQPFPHSAPALVDWQSNNTFASCTDWCIVHCKLGDRIPIKFOGHTNEV 352
Db 478 WDASTGCTGQFAFHAPALVDWQNTQAFASCTDQRHVCRLGWNPIKTFKGTNEV 537
QY 353 NAIKWPTGNLLASCSDMTLKIWSKQDNCVHDLQOHNKIYTIKWSPTGPTNNPNAN 412
Db 538 NAIKWCFQGLLASCSDMTLKIWSKQDNCVHDLQOHNKIYTIKWSPTGPTNNPNAN 597
QY 413 LMLASAFSTVRLWVDRCICHTITLKQEPYYSVAFSPDGYLASGFDKCVHIWNTQ 472
Db 598 LILASAFSTVRLWVDVRSCTIHTKTEPYYSVAFSPDGKHLASGFDKCVHIWSTQ 657

RESULT 10
Q9FN19 PRELIMINARY; PRT; 613 AA.
AC Q9FN19;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K8K14 (AT5967320/K8K14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned pl clones.";
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.W., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan W.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057698; AAL15328.1; -
DR EMBL; AY143932; AAN28871.1; -
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 2.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.

```

```

DR PROSITE; PS50896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 613 AA; 69772 MW; 04F40772311C0E76 CRC64;

Query Match
Best Local Similarity 46.0%; Score 1251; DB 10; Length 613;
Matches 248; Conservative 85; Mismatches 131; Indels 108; Gaps 7;

QY 2 SISDEVNLFVYRLOESGFSHSAPTFGIKSHISCSNINGALVPPAALISIQKGLQVFE 61
Db 3 SLTSELNLFVRYLQESGFTAAFTLGYEAGINKSNIDGNMVPFGALIKFVQKGLQVFE 62
QY 62 -----AEVSINEDGTLFDGRPIESLSLIDAVMPDVVTFQAVRDKLACQAAAAA 114
Db 63 MEANLSNSEDVIDEDFSFF--QPLDLISKVKELODMLEKKRKERDEKEDSKENDK 120
QY 115 AAAASQGSASAKNGENTANGE-----ENGAAHTIAN 143
Db 121 GVEREHEGDRNRAKEKDRHOKERERERERERERERERERERERERERERERER 180
QY 144 NHTDMVEVD-----GDVE----- 156
Db 181 REKORLKEKEEREIERERERERERERERERERERERERERERERERERERERER 240
QY 157 -----IPENKAVLRGHESEVFI CAMNPVSDLLASGSDSTARIWNLSENSTS 204
Db 241 DIVMTPTSTSHIPNSDVRILLEHTSEVCACAWSPSALLASGSDATARIWISPEGSFK 300
QY 205 -----GSTQLVLRHCIREGQDVPSNKDVTSLDWNSEGTLLATGSDGFRARIWTKGNL 258
Db 301 AVHTGRNINAILKHA---KGSNEKSKDVTLLDWNSEGTLLATGSDGQARIWTLNGEL 357
QY 259 ASTLQGHKGPIFALKWKNKGFIILSAGVDKTTIWDATHTGEAKQPFPHSAPALVDWQ 318
Db 358 1STLSKHKGPIFSLKWNKKGDIYLLTGSVDRTAVVDVKAEWKKQPFHSGPTLDVDRN 417
QY 319 NNTFASCTDMCIHVCKLGDRIPIKTFQGHNTNEVNAIKWDPTGNLLASCSDMTLKIWSM 378
Db 418 NVSFATSTSDSMIYCKIGETRPARTFTGHQGEVNCVKWDPTGSLASCSDSTAKIWN 477
QY 379 KODNCVHDLQOHNKIYTIKWSPTGPTNNPNANMLASAFSTVRLWVDVRCICHTL 438
Db 478 KQSTFVHDLRHTKTIYTIIRWSPTGPTNNPNKQTLTASASFDSTVRLWDAELGKMLCSF 537
QY 439 TKHQEPYYSVAFSPDGYLASGFDKCVHIWN 470
Db 538 NGHREPVYSLAFSPNGEYIASGSLDKSIHWS 569

RESULT 11
Q9SRJ9 PRELIMINARY; PRT; 524 AA.
ID Q9SRJ9;
AC Q9SRJ9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE LD24373P.
GN EBI OR CG4063.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brockstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

```

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY061326; AAL28874.1; -;
 DR FlyBase; FBgn0023444; ebi.
 DR GO; GO:0000074; P:regulation of cell cycle; IMP.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 3.
 DR PRINTS; PR000018; WD40; 4.
 DR PROSITE; PS000037; MYB_1; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 2.
 DR PROSITE; PS00082; WD_REPEATS_2; 3.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 524 AA; 52529 MW; 6C3D86110BA18D65 CRC64;

Query Match 34.0%; Score 923.5; DB 5; Length 524;
 Best Local Similarity 40.5%; Pred. No. 2.8e-57;
 Matches 218; Conservative 39; Mismatches 64; Indels 217; Gaps 12;
 QY 1 MSISSEVNFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIKQLQYV 60
 DB 1 MSFSSDEVNFVRYLQESGFLSHAYVFGHSHISQSNINGALVPPAALITLQKGLLT 60
 QY 61 EAEVSNEDGTLFDGRPIESLSIDAMPDV-----VOTRQ----- 96
 DB 61 EVSVSGDGEV--ARPIEGLSLIDAVMPVKPLKTVKTEPKGPAVDSSAPAGGNQNN 118
 QY 97 -----QAYRDKLA-----QQQAAAAAASQGSAAKNGENT 130
 DB 119 NAKPEIKIEPTGVAGSGAGNKIAGTTGTSTPTDQASAEVDSGNAANNAGTYAGNG 178
 QY 131 ANGEAE----- 135
 DB 179 AGNQAGTGGGNSSTPAGGLAAPGASQKQNSNEAGSSSGNAGNANATSTDDAASS 238
 QY 136 ---NG-----AHTANHTDM----- 148
 DB 239 TSTNGNSTSSVQPTSGLTGAGTGTSTNPDAAASGGASTATGSKAPSGAVTIIVGAQ 298
 QY 149 -----MEVDGDEIIPNKAIVLRGHESEV 172
 DB 299 GNVQSSSSNAQSPAGTSSSTSGAGTPAALVPMIDENIEIPSKARVLRGHESEV 358
 QY 173 PICAWNPVSDLLASGSGDSTARINLNSENSTSGTQLVLRHCHREGQDVPSNKNVTSLD 232
 DB 359 PICAWNPVSDLLASGSGDSTARINLWMSDANTN-SNQLVLRHCHIKGGAEPVSNKNVTSLD 417
 QY 233 WNSGCTLLATGSGYGEARIWTKDGNLSTLGQHKGPFFALKWKNKGNFILSAGVDKTTII 292
 DB 418 WNCDSLLATGSGYGEARIWTKDGNLSTLGQHKGPFFALKWKNKGNFILSAGVDKTTII 477
 QY 293 WDAHTGEAKQOFPFHSAPALVDVQSNNTFASC-----STDMCIHV-CK 335
 DB 478 WDASTG-----PMHPAICL-----SQCSSLGLGLADKPGCLLIQYGSADTCVPVGVCK 524

RESULT 12
 Q8VEG3 PRELIMINARY; PRT; 201 AA.
 AC Q8VEG3;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Similar to IRAP protein (fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC018512; AAH18512.1; -;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR PRINTS; PR00320; GPROTEINERPT.
 DR PRODOM; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 4.
 DR PROSITE; PS00678; WD_REPEATS_1; 2.
 DR PROSITE; PS00082; WD_REPEATS_2; 3.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT NON_TER
 SQ SEQUENCE 201 AA; 22038 MW; 6B945F137B491818 CRC64;

Query Match 32.5%; Score 882; DB 11; Length 201;
 Best Local Similarity 99.4%; Pred. No. 6.5e-55;
 Matches 158; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 314 VDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTL 373
 DB 1 VDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPTGNLLASCSDDMTL 60
 QY 374 KIWMSKQDNCVHDHLCQHNKEIYTIKWSPTGPTGNTNPNANLMLASASFDSTVRLWDVDRGI 433
 DB 61 KIWMSKQDNCVHDHLCQHNKEIYTIKWSPTGPTGNTNPNANLMLASASFDSTVRLWDVDRGI 120
 QY 434 CIHLTKHQEPVYSVAFSPDGRYLASGDFKCVHIWNTQ 472
 DB 121 CIHLTKHQEPVYSVAFSPDGRYLASGDFKCVHIWNTQ 159

RESULT 13
 Q8ZOR1 PRELIMINARY; PRT; 1227 AA.
 AC Q8ZOR1;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE WD-40 repeat protein.
 GN ALR0029.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AF003581; BAB77553.1; -;
 DR PIR; AE1810; AE1810.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003793; F:defense/immunity protein activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006915; P:apoptosis; IEA.
 DR GO; GO:0006952; P:defense response; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR00767; Disease_resist.
 DR InterPro; IPR002197; HTH_Fis.
 DR InterPro; IPR00792; HTH_LuxR.
 DR InterPro; IPR002182; NB-ARC.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00931; NB-ARC; 1.


```

DR Pfam: PF00400; WD40; 14.
DR PRINTS; PRO0364; DISEASERISIT.
DR PRINTS; PRO0320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR TIGRFAMs; TIGR01199; HTH_fis; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 12.
DR PROSITE; PS00082; WD_REPEATS_2; 14.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Complete proteome.
SQ SEQUENCE 1227 AA; 137236 MW; 466726939ED4F6F CRC64;

Query Match
Best Local Similarity 29.2%; Pred. No. 9.9e-23;
Matches 130; Conservative 73; Mismatches 171; Indels 71; Gaps 15;

QY 64 VSNEDG-TLFDGRPIESLIDAVMPDVVQTRQ-----QAYEDKLAQQQAAAAA 117
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 780 VSFSPDQQLASSGEDSTVRLWD-----VKTQCQWQIFEGHKKVYSVRFSPDGGOTLAS 833
QY 118 ASQQGSAK-----NGE--NTANGEENGAAHTIANNHDTMMVEVDG-----DV 155
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 834 CGEDRSIKLWDIQRGECVNTLWCHSSQVWAIA-----FSPDGRTLISGSDPDQARLWDV 887
QY 156 EIPPNKAVLGRHSEVFICAWNPVSDLLASGSDSTARIWNLSNSTSGSLQLVLRHCI 215
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 888 -ITGNSLNLGRYTRDYSVAFSPDSQILASGRDDYITIGLWNLKTGEC-----HPL 937
QY 216 REGGDVPSNKDVTSLDWNSEGLLATGSDYDGFARIW---TKDGNLSTLGHKGPIFA 271
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 938 R-GHQG-----RIRSVAFPDGKILASGSADNTIKLWDISDTHSKYITLITGHTNWWT 991
QY 272 LKNNKGNFILSAGVDTKIIDAHTGEAKQPPFHSAPALDVDMQ--SNNTFASGSDTMC 330
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 992 VVFPDPKHTLASSSEDRTIRLWDKDCGLQKLGHSWVWTVAFSPDGRILASGSADSE 1051
QY 331 IHVCKLGQDRPIKTFQGHTEVNAIKWDPKGNLLASCSDDMTLKIWSMKQDNCVHDLQGH 390
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1052 IKIWDVASKKCLQTLTDPOGMVSWAFSLDGLTLLASASEDQVTKLWNLKTGECVHTLKGH 1111
QY 391 NKEIYTIKWSPTGPTGNPNANMLASAFDSTVRLWDVDRGICHTLTK-HQEPVYSVA 449
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1112 EKQVSVAFSPNQ-----IAASGEDTIVKLWDISTGSCVDTLKHGHTAARISVA 1162
QY 450 FSPDGRYLASGDFKCVHINWTVQC 474
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1163 FSPDGRLLASGEDEXIQLWDMQNC 1187

RESULT 14
Q8X1P3 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P3
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C*40.
GN HET-E.
OS Podospira anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospira.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323584; AAL37300.1; -.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.

RESULT 15
Q8X1P5 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P5
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C.
GN HET-E.
OS Podospira anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospira.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323582; AAL37298.1; -.
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:43 ; Search time 13.3333 Seconds
(without alignments)
3708.183 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 2731

Sequence: 1 MSISSEVNFVLYRQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Piri:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501.5	18.4	535	2	S48268	probable membrane
2	481	17.6	1526	2	AC2239	WD-40 repeat prote
3	462	16.9	1356	2	T18521	beta transducin-li
4	459.5	16.8	1258	2	A12155	WD-repeat protein
5	451	16.5	1683	2	AF2071	WD-40 repeat prote
6	437	16.0	1227	2	AE1810	WD-40 repeat prote
7	436.5	16.0	1189	2	A12493	WD-repeat protein
8	436.5	16.0	1708	2	AE1866	WD-40 repeat prote
9	415	15.2	1711	2	AD1842	WD-40 repeat prote
10	404.5	14.8	1747	2	AC1842	WD-40 repeat prote
11	403.5	14.8	934	2	AG1889	WD-40 repeat prote
12	391.5	14.3	564	2	T40883	WD repeat protein
13	390.5	14.3	1693	2	S76086	beta transducin-li
14	385.5	14.1	676	2	AE2195	hypothetical prote
15	385	14.1	1551	2	AB2410	WD-repeat protein
16	374.5	13.7	515	2	S19487	hypothetical prote
17	373	13.7	1189	2	AE2154	WD-repeat protein
18	369.5	13.5	502	2	T41148	trp-asp repeat con
19	366.5	13.4	304	2	AG1837	WD-40 repeat prote
20	362	13.4	265	2	AE1890	WD-repeat protein
21	358	13.1	437	2	S05357	hypothetical prote
22	352.5	12.9	677	2	AE1861	serine/threonine k
23	344.5	12.6	786	2	AG2375	WD-40 repeat-prote
24	340	12.4	333	2	G85034	probable WD-repeat
25	339.5	12.4	559	2	AB2202	hypothetical prote
26	334.5	12.2	410	2	S48052	platelet-activatin
27	331	12.1	323	2	T02617	hypothetical prote
28	331	12.1	777	2	T41075	hypothetical WD-re
29	329.5	12.1	409	2	S36113	Lis-1 protein - hu

ALIGNMENTS

RESULT 1

S48268 probable membrane protein YBR103w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0832

C:Species: Saccharomyces cerevisiae

C>Date: 01-Aug-1995 #sequence revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S48268; S45971; S44683

R:Manhaupt, G.; Stucka, R.; Ehmlé, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48268

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-535 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAAS5606.1; PID:g476059

R:Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45927

A:Accession: S45971

A:Molecule type: DNA

A:Residues: 1-535 <FE2>

A:Cross-references: EMBL:Z35972; MIPS:YBR103w

C:Genetics:

A:Gene: SGD:SIF2

A:Cross-references: SGD:S0000307

A:Map position: 2R

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

C:Keywords: transmembrane protein

F:355-388/Domain: WD repeat homology <WD1>

F:397-429/Domain: WD repeat homology <WD2>

F:434-450/Domain: transmembrane #status predicted <TMM>

Query Match

Best Local Similarity 18.4%; Score 501.5; DB 2; Length 535;

Matches 143; Conservative 101; Mismatches 239; Indels 79; Gaps 16;

Qy 1 MSISSEVNFVLYRQESGFSHSAFTFGIESHSQSNN-GALVPPAALISIIQKQLOY 59

Db 1 MSITSEELNYLWKYQCEMGHEVSALALQDTRVLEFDEKYEHIPIGLTIVNLVQRGILY 60

Qy 60 VEAERVSINEDG---TLFDGRPIESLSLIIDAVMPDVVQTRQQAYRDKLAQHAAAAAAA 116

Db 61 TESELMVDSKGDTSALNEHHLSDFNLVQALQID-----KERFPE--ISSEGRFTL 109

Qy 117 ATNQCSAKNGENTANGEANGAHTIANNHTDMV-EVDGVEIPEPSKAVLVRGHESVFFIC 175

Db 110 ETNSE-SNKAGEDGASTVERETQEDDTNSIDSSDDLDFGVKLI--LKEIV---KLDNIVSS 163

Qy 176 ANPVPD-LIVSGSGDSTARIWNLSNSTSGTPQ-----LVLRHCIRGGQVPSNKDV 228

Db 164 TWPPLDESILAYGEKNSVARLARIIVETDQEGKKYKWLITIAELRHPFALSASSGKTTNQV 223

PF20 protein, micr

WD-repeat protein

hypothetical prote

TATA box-binding p

hypothetical prote

hypothetical prote

WD-repeat protein

MET30 protein - ye

WD-40 repeat prote

hypothetical prote

beta transducin-li

WD-40 repeat regul

beta transducin-li

transcription init

hypothetical prote

QY 229 TSLDWNSEGTLLATSGYDGFARIWTKDGNLSTLQHKGPFPALKWKNKGNFILLSAGYDK 288
DB 224 TGLAWSHDGNISVTGVEGELRLNKTGALLVNLPHFRAPIVSVRWKNDGTHIISMDEVN 283
QY 289 TTIINDAHTGEAKQOPFF-----HSAP---ALDWDQSNNTFASCTDTCMHV 333
DB 284 VTILNVIISGTVMQHPKELKTCGSSINAENHSGDGLGVDEWDDKFPVPGKGAIFV 343
QY 334 CKLQDQRPKTFQGHTEVNAIKWPTGNLLASCDDMTLKIWSMKQNCVHDLOAHNKE 393
DB 344 YQITKTPTKLIGHGHPISVLEFNDTNKLLSASDDGLRIWHGGNGNSQNSFYGHQS 403
QY 394 IYTIKSPGPGTGNPNANMLASASFDSTVRLWDVDRGICHTLTKHQPYPYVAFSPD 453
DB 404 IYASWV-----GDKVISCMDGSLVRLSLKQNTLLASIVDGVPIFAGRISQD 453
QY 454 GRYLASGSPDKCVHIW-----NTQTAL-----VHSYRGTGGIFEV 490
DB 454 GQYAVAFMDGVVYDLKLNKSKRSRSLYGNRDGILNPLPIFYASYOSSQNDYIFDL 513
QY 491 WNAAGDKVGA--SASDGSNCVL 510
DB 514 WNCAGNKRISVAYSLQEGSVVAI 535

RESULT 2
AC2239
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2239
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2239
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE75165.1; PID:g17132599; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3466

Query Match 17.6%; Score 481; DB 2; Length 1526;
Best Local Similarity 30.3%; Pred. No. 1.7e-26;
Matches 118; Conservative 77; Mismatches 149; Indels 46; Gaps 11;
QY 136 NGAHTIANNHTDMVEVDGVEIPSNKAV-VLRGHESEVIFICAWNPVSDLLVSGSDSTAR 194
DB 1127 NGV-TLANGSSDQIVRLWD--ISSKKLYTQHTNWNVAVAFSPDGATLASGSGDQTVR 1183
QY 195 IWNLSNSTSGPTQLVLRHC--IREGQDVPNSKDVTSLDWNSGHTLLATGSDGFARIV 252
DB 1184 LWDISSK-----CLYLQ-----HTSWNSVVFNPDPGSLASGSDQTVRLW 1227
QY 253 -TKDGNLASTLQHKGPFPALKWKNKGNFILLSAGVDTKTIINDAHTGEAKQOPFFHSA 311
DB 1228 EINSKCLCTFQHTSWNSVWFNPDPGSLASGSDQTVRLWDLISSKCLHTFQHT--- 1284
QY 312 LDVDWQSNNTF-----ASCSTDMCHVCKLQDQRPKTFQGHTEVNAIKWPTGNLL 364
DB 1285 ---NWNVSVAFNPDPGSLASGSDQTVRLWEISSKCLHTFQHTSWNSVVFSPDGTWL 1341
QY 365 ASCSDDMTLKIWSMKQNCVHDLOAHNKEIYTIKWSPTGPTGNPNANMLASASPDSTV 424
DB 1342 ASGSDQTVRLWISISSGCLYTLFHTNWNVGSVIFSPDG-----AILASGSGDQTV 1392
QY 425 RLWDVDRGICHTLTKHQPYPYVAFSPDGRYLASGDFKCVHNTQTGALVHSYRG-T 483

DB 1393 RLWSISSGKLYTQGHNNVWGSIVFSPDGTLLASGDDQTVRLMNNISSGCLYTLHGHI 1452
QY 484 GGIFEVCMNAAGDKVGSASDGSVCVLDLR 513
DB 1453 NSVRSVAFSSDGLILASGSDDETIKLDVYK 1482

RESULT 3
T18521
beta transducin-like protein - Podospora anserina
C:Species: Podospora anserina
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18521
R:Saube, S.; Turcq, B.; Begueret, J.
Gene 162, 135-139, 1995
A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
A:Reference number: Z18944; MUID:96009891; PMID:7557402
A:Accession: T18521
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1356 <SAU>
A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1
C:Genetics:
A:Gene: het-e1
A:Introns: 761/3

Query Match 16.9%; Score 462; DB 2; Length 1356;
Best Local Similarity 27.3%; Pred. No. 3.4e-25;
Matches 121; Conservative 74; Mismatches 181; Indels 68; Gaps 11;
QY 78 IESLSLIDAVMPDVVQROQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 130
DB 820 ISTISVVEAEWNACTQT-----LEHGSSVLSVAFSADGQVRSAGSDDKTIKWD 870
QY 131 ANGENGAHTIANNHTDMVEVDGVEIPSNKAVVLRGHESEVIFICAWNPVSDLLVSGSD 190
DB 871 ASG--TGTQT-----LEHGSSVLSVAFSADGQVRSAGSDDKTIKWD 904
QY 191 STARIWNLSENSTSGPTQLVLRHCIREGQDVPNSKDVTSLDWNSGHTLLATGSDGFAR 250
DB 905 KTIKIWDASGTC---TQTLGH---GGR-----VQSVAFSPDQVRSAGSDDKTIK 950
QY 251 IW-TKDNGLASTLQHKGPFPALKWKNKGNFILLSAGVDTKTIINDAHTGEAKQOPFFHSA 309
DB 951 IWDASGTCCTQTLLEGHSSVLSVAFSPDQVRSAGSDDKTIKWDASGTCCTQTLGHGG 1010
QY 310 PALVDWQ--SNTPASCTDMCHVCKLQDQRPKTFQGHTEVNAIKWPTGNLLASCS 368
DB 1011 SWSVAFSPDQVRSAGSDDKTIKWDASGTCCTQTLGHGGVQSVVFPDQVRSAGS 1070
QY 369 DDMTLKIWSMKQNCVHDLOAHNKEIYTIKWSPTGPTGNPNANMLASASPDSTVRLWD 428
DB 1071 DDHTIKIWDASGTCCTQTLLEGHSSVLSVAFSPDQVRSAGSDDKTIKWD 1121
QY 429 VDRGICHTLTKHQPYPYVAFSPDGRYLASGDFKCVHNTQTGALVHSYRG-T 487
DB 1122 AASGTCCTQTLLEGHGGVHVSFAVSPDQVRSAGSDDKTIKWDASGTCCTQTLGHGGVQ 1181
QY 488 EVCWNAAGDKVGSASDGSVCVLD 511
DB 1182 SVAFSPDQVRSAGSDDKTIKWD 1205

RESULT 4
A12155
WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: A12155
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001

Db 1164 SPDGRLLASGEDEXIQLWDMQ 1185

RESULT 7
AI2493

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF2493
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF2493
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1189 <KUR>
 A:Cross-references: GB:BA000020; PIDN:BA78213.1; PID:g17135667; GSPDB:GN00180
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr7129
 A:Genome: plasmid

Query Match 16.0%; Score 436.5; DB 2; Length 1189;
 Best Local Similarity 28.7%; Pred. No. 2e-23;
 Matches 108; Conservative 65; Mismatches 142; Indels 61; Gaps 9;

QY 167 GHSEVFICAWNPVSDLLVSGSDSTARIWNLSE-----NSTSGPTQLVLRHCIRGGQ-- 220
 DB 686 GHDACVMSVVFHPVQGLATAGEDNTIKLWELQSGCCKTLQGHQHWVWTFIAFNSGGRIIL 745

QY 221 -----DVPSENK-----DVTSLDNSEGLTLLATSGYDFARIW-TKQDN 257
 DB 746 ASGSPQNVKLDIHTGKCVMTLQHTGVTVSVAPNPKDNLGSGYDQSVKVMKRTGR 805

QY 258 LASTLQGHKGPFPALKKWKNGNFIKILSAGVDTIWDANTGEAKQFPFHSAPALDV--D 315
 DB 806 CLDTLKKHTNRIWSVAFHPQGLHFLVSGDDHAAKIWEIQTGQCIKTFOGHSNATYIAHN 865

QY 316 WQSNNTFASCTDWCIVHVKLG-----QDRPKTFQGHNTNEVNAIKWDPGTGNLLAS 366
 DB 866 WE-HSLLAGSHEDQTKLWDLNLHSPKSNVNTHPRIILQHSNRVSVFVSTGQLLAS 924

QY 367 CSDMTLKIWSKQDNCVHDLOAHNKEIYTIKSPGTGPTNPNANMLASFPSTVRL 426
 DB 925 GSADRTIKLWSPHTGQCLTLHGSGWMAIAFSLD-----DKLLASGSYDHTVKI 975

QY 427 WDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSPDKCVHIWNTOTGALVHSYRGTGI 486
 DB 976 WDVSSGQCLTLQGHFGSVLAVAFSCDKLPSFGYKELVKQWDETGYCIQT----- 1028

QY 487 FEVCWNAAGDKVGASA 502
 DB 1029 -----WEADSNRVNVA 1040

RESULT 8
 AE1866
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AE1866
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE1866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1708 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA72436.1; PID:g17129823; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0478

Query Match 16.0%; Score 436.5; DB 2; Length 1708;
 Best Local Similarity 30.5%; Pred. No. 3.3e-23;
 Matches 123; Conservative 70; Mismatches 169; Indels 41; Gaps 15;

QY 119 NQGSAGKNGTANGBENGAAHTIA---NNHT-DMMEVDGDEIPEPSNKAVV--LRGHESE 171
 DB 1298 NRQG---NLKTLISGHTAGVATVPFNGETIGSASIDATLKLWSPQGLLGLTKGHSW 1354

QY 172 VFICAMNPVSDLLVSGSDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDVTS 231
 DB 1355 VNSVSPSPDGRIFASGSRDKTVTLRWDE-----VLLRNPKGDG-----NDWVTSI 1400

QY 232 DWNSEGLTLLATSGYDFARIWTKDGNLSTLQGHKGPFPALKKWKNGNFIKILSAGVDTI 291
 DB 1401 SFSGDTTAAASRDQTVKILSRHGLKLLNTFKGTGSIWGVAPNRPWIASAKDQTVK 1460

QY 292 IWDATGEAKQFPFHSAPALDVWQSN--TFASCSTDMCIHVCKLGQDRPKITFGHTN 350
 DB 1461 LMW-QDGKILHTLQGHQDAVLAVANSSDQSVIASAGDKIVKWSQG-QGLLHTLQHTD 1518

QY 351 EVNAIKWDPGTGNLLASCDSDMTLKIWSKQDNCVHDLOAHNKEIYTIKSPGTGPTNPN 410
 DB 1519 AVNWVSFPDGKILASVSDDTTVKLMS-RDQQLLHTLKEHSRRVNGVAVSPDQG-- 1571

QY 411 ANMLASASPDSTVRLWVDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSPDKCVHIW 470
 DB 1572 ---ILASASIDGTVKLWNRD-GSLSRNLPGDGSFISVSPDGKMLAANSDDQ-IRLMN 1626

QY 471 TQTGALVHSYRG-TGGIFECVWNAAGDKVGASASDQSVCLDL 512
 DB 1627 -QKGTLLMWLKGDKDELTSVTFSPDSQILAVGGGNGKVFIFNL 1668

RESULT 9
 AD1842
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD1842
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD1842
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1711 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA77808.1; PID:g17135262; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0284

Query Match 15.2%; Score 415; DB 2; Length 1711;
 Best Local Similarity 24.9%; Pred. No. 1.2e-21;
 Matches 123; Conservative 75; Mismatches 180; Indels 116; Gaps 14;

QY 65 SINEDGTL---FDRPIESLISLIDAVMPDVVQTRQAYRDKLAQQAHAHAATAATNQ 120
 DB 1246 SASDDGIRLWSLGRPLITI-----PSHTKQVLAVTFSPDG 1282

QY 121 QGSAGKNGTANGBENGAAHTIANHTDMMEVDGDEIPEPSNKAVVLRGHESEVFCANPV 180
 DB 1283 QTVSAG-----ADNTVKLWSRNGTL-----LTLGHNNAVQVIFSPD 1322

QY 181 SDLLVSGSDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDVTSLDNSEGTL 240
 DB 1323 GRLTASADKTIILSRDGN-----IL-----GTFAGHHEVNSLSFSPDGNIL 1367

QY 241 ATGSYDGFARITWTKDGNLSTLQGHKGPFPALKKWKNGNFIKILSAGVDTIINDA----- 295

```

Db 1368 ASGSDNTVRLWTVNRTLPKTFYGHKGVSYSYRFPNDGKKTISLSTDSTMTKWSLDGKLL 1427
QY 296 -----HT-----GEAKQFPFHSAAPALDWDQSN- 319
Db 1428 QTLSSPLPDVTSISFTPDNKIVALASPDHTIHLNRYQGLLRSLPCHNHWTLSLSPNK 1487
QY 320 NTFASCSSTDMCHVCKLGDORIPKTFQGHTEVNEVNAIKWDPTGNLLASCSDDMTLKISWK 379
Db 1488 QILASGSADTKIKLSV-NRGLLKTLHGNGWVTDIKFSADGKNIVSASADTKIKLSL- 1545
QY 380 QDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAPDSTVRLWDVDRGICIHILT 439
Db 1546 DGLRTLTLOHGSASVSNLSPDQ-----TLASTQDETILKNLW-GEIYILR 1595
QY 440 KIQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTALVHSYRG-TGGIFEVCWNAAGDKV 498
Db 1596 GHSDDVYNLSFSPDGTITASDDGTIKLWNVENGTLTKTQCHRGVRSVSFSPDGKIL 1655
QY 499 GASASDGSVCVLDL 512
Db 1656 ASGGHDTTVKWNL 1669

RESULT 10
AC1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1747 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:gl7135261; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0283

Query Match 14.8%; Score 404.5; DB 2; Length 1747;
Best Local Similarity 26.9%; Pred. No. 7.2e-21;
Matches 112; Conservative 67; Mismatches 158; Indels 79; Gaps 10;
QY 165 LRHSESEFICAWNPVSDLLVSGSGDSTARIWNL-----SENSTSGPTQ 208
Db 1310 LTGHRKRTSVAFSPDGKILASAGDKTIKFNTDGFELKTIHAHQVNSINFSSDKT 1369
QY 209 LV-----LRHCIREGGQDV-----PSNK----- 226
Db 1370 LVASGADSTMKWKIDGLTIKITSRGQIRDVTFSPDNKVIASASSDKTVRIQLNVQK 1429
QY 227 ----DVTSLDNSEGLTATSGYDFARIWTKDGNLA----STLQCHKGPFAKWNKKG 278
Db 1430 SOKSNVSVSNFNPDKTFASAGWGNITIQRE-FLAHSLSLTQKNQNIITVSYSYSPDG 1488
QY 279 NFILSAGVDKTTIINDAHTGEAKQFPFHSAAPALDWDQ--SNNTFASCSSTDMCHVCKLG 337
Db 1489 KTIATASADNTIKLWDSQTQQLIKTLTGHKORITTLTSFHPNQTIASGADKTIKIWRN 1548
QY 338 QDRPIKTFQGHTEVNEVNAIKWDPTGNLLASCSDDMTLKISWKQDNCVHDLQAHNKEIYTI 397
Db 1549 DQQLRTLTGHNDVTSYVNFSPDGGFLASGSDTNVTKW-QTDGELIKNIITGHGLATSV 1607
QY 398 KWSPTGPTNNPNANMLASASFDSTVRLWDVDRGICIHILTQKQEPVYSVAFSPDGRYL 457
Db 1608 KFSPD-----SHTLASASWNTIKLWQVTDGKLIINLNGHIDGVITLSLFSFDPGEIL 1658
QY 458 ASGSDKCVHIWNTQTALVHSYRG-TGGIFEVCWNAAGDKV GASASDGSVCVLDL 512

```

```

Db 1659 ASGSADNTIKLWNLPNATLLKTLGHGKINTLAFSPDGKTLTLLSGGEDAGVWVWNL 1714

RESULT 11
AG1889
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1889
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-934 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072622.1; PID:gl7130010; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0664

Query Match 14.8%; Score 403.5; DB 2; Length 934;
Best Local Similarity 23.1%; Pred. No. 3.5e-21;
Matches 128; Conservative 115; Mismatches 197; Indels 113; Gaps 21;
QY 8 VNFVLYRYLQESGFSHSAFTFGTIESIQSNGINGALVPPAALISIIQKQYVEAEVYSIN 67
Db 231 ISSLVTLGIIALGLAGVANTQQKARNSETK-----AISSAESFLNANLEF- 277
QY 68 EDGLTFD---GRPTESLSDLA-VMPDVVQTRQAY-----RDKLACQAAAAAATAAAT 118
Db 278 -DGLIASIRAGRRIKGTGDIDANTRTQITETLQOSINFVEKRNLAHDGMLSESVSFSD 336
QY 119 NQGSASAKNGENTANGEGAHTIANNHTDMVEDGDVEIPSNKAVLVRGHSEVFI-ICAM 177
Db 337 SKFIATASRQTKV-----IWSLDGKKQL-----VVLREKGGEGFNSVAF 376
QY 178 NPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGO--DVPSNKD-VTSLDMN 234
Db 377 SPDGTLMATGSDWNTAKIWS-----REGKRLHTLDGKHAIVLEAFS 418
QY 235 SEGTLIATGSDYGFARIWTKDGNLASTLGOHKGPFAKWNKGNFILSAGVDKTIWID 294
Db 419 PDSQLIATASDNTVTKLWSREGKLLHTLEGHKDKVNSITFSPDGQLIATVGNWNTMKLWN 478
QY 295 AHTGEAKQFPFHSAAPALDWDQ-----SNNTFASCSSTDMCHVCKLQDORPIKTFQGH 349
Db 479 L-DGKELRTFRGHQ----DMIWSVSFSPDGKQIATASGDRTVKLWSL-DGKELQTLRGHQ 532
QY 350 NEVNAIKWDPTGNLLASCSDDMTLKISWKQDNCVHDLQAHNKEIYTIKWSPTGPG---- 405
Db 533 NGVNSVTFSPDGKLIATASGDRTVKLWNSKGQF-LETLYGHTDAVNSVAFSPDGTSIATA 591
QY 406 -----TNNPNANLM-----LASASPDSTVRLWDV--DRGICI 435
Db 592 GNDKTAIXKLNLSNLSIIVRGHEDEVFDLVFSPNGKYIATASWDKTAKLMSIVGDKLQEL 651
QY 436 HTLTKHQEPVYSVAFSPDGRYLAGSFDKCVHIWNTQTALVHSYRG-TGGIFEVCWNA 494
Db 652 RTFNGHQGRVNLKLSFSPDGKYIATTSWDKTAKLWNL-DGLQKTTLGKDKTVMSVNFSPD 710
QY 495 GDKVGASASDGSV 507
Db 711 GLIATASDEKTV 723

RESULT 12
T40883
WD repeat protein - fission yeast (Schizosaccharomyces pombe)

```

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40883
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21954
A:Accession: T40883
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-564 <WO>
A:Cross-references: EMBL:AL031764; PIDN:CAA21113.1; GSPDB:GN00068; SPDB:SPCC1235.09
A:Experimental source: strain 972h-; cosmid c1235
C:Genetics:
A:Gene: SPDB:SPCC1235.09
A:Map position: 3
A:Introns: 18/1; 273/3; 413/3

Query Match 14.3%; Score 391.5; DB 2; Length 564;
Best Local Similarity 24.7%; Pred. No. 1.3e-20;
Matches 134; Conservative 106; Mismatches 201; Indels 101; Gaps 24;

QY 3 ISSDEYNFLVRYLQSGFSAFTGIBSHISQSNIN--GALVPPAALISIIQKGLQY 59
DB 1 MDINQVNIWRYLKEGYSHTKFAFERETGI--QNLDRKQWGSTCQGVGVILQKGLQY 58
QY 60 VEAEV-----SINEDG--TLFDGRPIESLSLD-----AVMPDVQV----- 94
DB 59 VELEKHYVNDHNSNEEASKTSIDGE-----SLVNEPCKLPYLIVPHICETTLTKADST 113
QY 95 -----RQAVRDKLQCHAAHAAAAAATNQGSAKNG-----ENTANGENGAAHTIA 142
DB 114 NGPCEHNSNDHQLKILQDKSGSPSPVMPFEDKIEKRDIITWADENVEKDPARPIA 173
QY 143 NNHTDMEVDGVEIISNAKAVLRGHE---SEVF-----ICA-WNPVSD-----LLV 185
DB 174 VYNSPV-----TEITEIKQVTFPTGGEDIKSDFFKVIPTKHPVTCADMRPLLEQNYHYVE 228
QY 186 SGSGDGTARINLNS-----ENSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWSNSETGLIA 241
DB 229 FSGTMTNATLASVISICEQNDFAKTD---YCL-----QSSFDNQDITGVANNNSGSFLA 280
QY 242 TGSYDGFARIWKDGNLSTLQGHKQPIFALKKNKGNFILSAGVDKXTIINDAHTGEAK 301
DB 281 YAPFSGVIEIYDGHGQILSLFHNNKGFVLSLKWSGTDTYLAAGSADGTTLPD---QLK 336
QY 302 Q-QFPPH--SAPALVDVQSNNTFASCSTDMCTHVCGLGQDRPIKTF-QCHTNEVNAIKW 357
DB 337 QTQISIDTLASSVLDIEWISFDEFVTSDEGSLRVYKVGKAPVSTVSHADNSIVALRY 396
QY 358 DPTGNLLASCDDMTLKINSMKQD---NCVHDLQAHNKEIYTIKWSPTGPTGNTNPNANILM 414
DB 397 NLRISLLLTASSDTTVKLNSRGDAGAFELH-VFSPSSPVCIDW-----NLREGTPI 448
QY 415 LASASFDSTVRLNDVDRGICHTLTKHQPVPVSVAFSPDGRVLASGPKCVHINNTQTG 474
DB 449 LAVAS-NSIVSMYNAISLQOLAVMRHTAPVSALSFHNGRYLATGDTSGGVCISCKTA 507
QY 475 AL 476
DB 508 KL 509

RESULT 13
S76086
beta transducin-like protein, 190K - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sl10163
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C:Accession: S76086
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76086
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1693 <KAN>
A:Cross-references: EMBL:D63999; GB:AB001339; NID:G1001396; PIDN:BAA10064.1; PID:di010711
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:1051-1084/Domain: WD repeat homology <WD01>
F:1092-1125/Domain: WD repeat homology <WD02>
F:1133-1166/Domain: WD repeat homology <WD03>
F:1174-1207/Domain: WD repeat homology <WD04>
F:1256-1289/Domain: WD repeat homology <WD05>
F:1397-1330/Domain: WD repeat homology <WD06>
F:1420-1453/Domain: WD repeat homology <WD07>
F:1461-1494/Domain: WD repeat homology <WD08>
F:1502-1535/Domain: WD repeat homology <WD09>
F:1584-1617/Domain: WD repeat homology <WD10>
F:1625-1658/Domain: WD repeat homology <WD11>
F:1625-1658/Domain: WD repeat homology <WD12>

Query Match 14.3%; Score 390.5; DB 2; Length 1693;
Best Local Similarity 28.5%; Pred. No. 7.2e-20;
Matches 107; Conservative 53; Mismatches 137; Indels 79; Gaps 10;

QY 100 RDKLAQCHAAHAAAAAATNQGSAXNGENTANGE--ENGAHTIANNTDMMEVDGDVEI 157
DB 1362 RDKTARLWTEGECVAVLADHQGVWREGQSPGQWIVTGS---ADKTAQLWNLVG--- 1414
QY 158 PSNKAVALRGRHSEVFTCAWNPVSDLLVSGSGSTARIWNLSNENSTSGPTQLVLRHCIRE 217
DB 1415 --KKLTVLRGHQDAVLNVRESPPSQXIVTASDKGTARVWN-----NTGRELAVALRH--- 1463
QY 218 GGQDVPSNKDVTSLDWNSEGTLLATGSDGCFARIWTKDGNLSTLQGHKQPIFALKKNKK 277
DB 1464 -----YKMIIFRAEFSADGQFIVTASDNTAGIWEIVGREVICRGHSPVYFAFSAD 1517
QY 278 GNFTLSAGVDKTTIINDAHTGEAKQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLG 337
DB 1518 SRVILTASVDNTARIWDF-----LG 1537
QY 338 QDRPIKTFQCHTNEVNAIKWDPGNTGNLLASCDDMTLKINSMKQDNCVHDLQAHNKEIYTI 397
DB 1538 --RELLTLAGHSQIVYQARFSPGEGNLTATVSADHTARLWD-RSGKTVAVLVYGHQGLVGTV 1594
QY 398 KWSPTGPTGNTNPNANILMASASFDSTVRLNDVDRGICHTLTKHQPVPVSVAFSPDGRYL 457
DB 1595 DWSPDGQ-----MLVTASNDGTARLWDLG-SRELLTLEGHGNVRSAPFSPDGRWV 1644
QY 458 ASGSFDCVCHINNTQT 473
DB 1645 LTSSADGTAKLWPKVT 1660

RESULT 14
AH2195
hypothetical protein alr3119 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2195
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2195
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-876 <KUP>
A:Cross-references: GB:BA000019; PIDN:BA874818.1; PID:gl7132214; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3119

Query Match 14.1%; Score 385.5; DB 2; Length 676;
Best Local Similarity 24.2%; Pred. No. 4.6e-20;
Matches 127; Conservative 82; Mismatches 191; Indels 125; Gaps 14;

QY 21 FSHSAFTFGIEHSHISOSNINGALVPPAALISIIKGLQYVEA-----EVS 65

DB 226 FSLGATCFHLLTGINSNL-----FVEQGYSWVESWQYWNSTNSDRNEGYL 273

QY 66 INEDGTLFCRPIESLSLIDAVMPDVVQROQAYRDK-----LAQOHAIAAAAAA 117

DB 274 VKVLNKLLETDIQRVQSADEWMDLTKQSLRLKTIPIKSAIFSSWSASTSLTAST 333

QY 118 TNOQG-SAXNGE-----NTAN-----GEENGAAH--TIANNHTDMMEVDGVEIPIKNAW 164

DB 334 TKQAWKLLNGRLKQOLLINTMSALLGLVGVLQSLPOLITKFSFI-----STQPYT 386

QY 165 LRGESEVFCIAWNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVP 224

DB 387 LKGHASDVNSVAFSPNGEFLAGSDDKTIKWNL----- 420

QY 225 NKDVTSLDWNSGTLATGSDYGFARIWTKDGNLASTLQHKXGPIFALKWNKGNFILSA 284

DB 421 -----KNKQKIHTLPCHSGWVAIAFSPDGKTLAST 451

QY 285 GYDKTIIWDAHTGEAKQOPPHSAPALDVQO-SNNTFASCTDMCIHVCKLGQDRPK 343

DB 452 GADKTIKLNLTATGEIRHLKHSQGVASVAFSPDGKTLASGLDKTIKLNPAATGKEIR 511

QY 344 TPQCHTNEYNALIKWDPNTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTG 403

DB 512 TI-QEHSSGVANVAFSPDGKTLASGSWDKTIKLNLTTSKVHTLKGSDLVWSVAFN--- 568

QY 404 PGTNNPNANLMLASPSFSTVRLWDVDRGICHTLTKHQPVPYVAFSP-DCRYLASGSF 462

DB 569 -----SDSQTLASGSKDKTIKMLNSTGKTIRTLRSHSDKVNVAIYPRDSTVLASGN 622

QY 463 DKCVHIWNTQTGALVHSY-RGTGGIPEVCWNAAGDKVGASASDGS 506

DB 623 DNTIKLWLTGTEIIRTLKRDSGYIYSIVSPDGNL---ASGGS 664

RESULT 15

AB2410

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AB2410

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; NUID:21595285; PMID:11759840

A:Accession: AB2410

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1551 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076533.1; PID:g17133971; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr4834

Query Match 14.1%; Score 385; DB 2; Length 1551;

Best Local Similarity 22.3%; Pred. No. 1.6e-19;

Matches 120; Conservative 97; Mismatches 192; Indels 130; Gaps 18;

QY 48 ALISIIQGLQYVEVNEINEDGTLFDGRPIESLSLIDAVMPDVVQROQAYRDKLAQOH 107

DB 903 ALMSAMRSQ----KALQALVKDGRSLAKYPATSPALLQTLTDNIQERNQ-----FQGH 952

Search completed: August 9, 2004, 16:48:51

Job time : 15.3333 secs

QY 108 AAAAAAAAAAATNOQGSAKNGENTANGENGHAHTIANHNTDMMEYDGVDEIPSNKAVVLRG 167

DB 953 QAWVRSVS-----FSRDGQYILTASDDCTARLWN-----LQKGKQLISLQG 992

QY 168 HSEVFFICAWNPVSDLLVSGSGDSTARIWNLSNSTS-----GPTQLVL-----RHCIR 216

DB 993 HEDTIANFSPDGKYIATASSDRTARLWNFSGQQLAKFOGHQGYVRSVSFSPDGKHIAT 1052

QY 217 EG-----GQDV-----PSNKDV-TSLD-----WNSEGTILL-- 240

DB 1053 AGDDHTARLWSFSQQLVQFPGHQGTVMCISFSFDGKHIAATAADDRIVRLWNLKGLVLR 1112

QY 241 -----ATGSDYGFARIWTKDGNLASTLQHKXGPIFALKWNKGNF 280

DB 1113 FPGHDCVWDVSPDSQYIATASSDGTSLRNLAGEQITFRGHQGVVWSVFSFSPNGQY 1172

QY 281 ILSAGVDKTIIWDAHTGEAKQOPPHSAPALDVQOSNNTF-ASCSTDMDCIHVCKLGQD 339

DB 1173 IATTSSTRTARVWNLN-GQQLAQFSGHODYVRSVSPDGKYIATASSDRTVRLWHLNKQ 1231

QY 340 RPIKTFQCHTNEYNALIKWDPNTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKW 399

DB 1232 Q-FSAFQGHQSTVRSVDFSPDGKQKVTAADDRIVRLWNLKGEELLQFL-GHRGKVMVSF 1289

QY 400 SPTGPGTNNPNANLMLASPSFSTVRLWDVDRGICHTLTKHQPVPYVAFSPDGRYLAS 459

DB 1290 SPDGK-----YIATTSSTRTVRLWDI-TQQLQQQFPGHGTVMWSVFSFSPDGKHIAT 1339

QY 460 GSFDPKCVHIWNTQTGALV-----HSYRGTGGIPEVCWNAAGDKVG 499

DB 1340 ASSDLTTRLWLSLDGQELMQFKGHDKWVRYVVSFCNGQHIATAADDDCTARLWNLAGRQVG 1398



GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2004, 16:41:13 ; Search time 34.6667 Seconds
(without alignments)
4678.164 Million cell updates/sec

Title: US-09-987-701-12
Perfect score: 2731
Sequence: 1 MSISSEVNFVLYRLQESG.....GDKVGASASDGSVCVLDLRK 514

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2731	100.0	514	11	Q9EQD4		Q9eqd4 mus musculus
2	2727	99.9	514	11	Q8BHJ5		Q8bhj5 mus musculus
3	2724	99.7	514	11	Q8CBG4		Q8cbg4 mus musculus
4	2645.5	96.9	519	13	Q7SZM9		Q7szm9 xenopus lae
5	2466.5	90.3	527	11	Q8BYQ4		Q8byq4 mus musculus
6	2462.5	90.2	527	11	Q8BMM0		Q8bmm0 mus musculus
7	2439	89.3	577	4	Q8EUJ2		Q8euj2 homo sapien
8	1956	71.6	412	11	Q8C0A1		Q8c0a1 mus musculus
9	1906.5	69.8	700	5	Q9XZK1		Q9xzk1 drosophila
10	1388	50.8	613	10	Q9FN19		Q9fn19 arabidopsis
11	1101	40.3	201	11	Q8VEG3		Q8veg3 mus musculus
12	920.5	33.7	524	5	Q8ERJ9		Q8erj9 drosophila
13	480	17.6	1356	3	Q8X1P4		Q8x1p4 podospora a
14	477	17.5	1356	3	Q8X1P5		Q8x1p5 podospora a
15	472	17.3	1356	3	Q8X1P3		Q8x1p3 podospora a
16	445	16.3	1376	3	Q8X1P2		Q8x1p2 podospora a

17	437	16.0	1227	16	Q8Z0R1		Q8z0r1 anabaena sp
18	436.5	16.0	1189	16	Q8YL09		Q8yl09 anabaena sp
19	436.5	16.0	1708	16	Q8YZ12		Q8yz12 anabaena sp
20	415	15.2	1711	16	Q8Z019		Q8z019 anabaena sp
21	404.5	14.8	1747	16	Q8Z020		Q8z020 anabaena sp
22	403.5	14.8	934	16	Q8YZ23		Q8yz23 anabaena sp
23	391.5	14.3	564	3	Q74845		Q74845 schizosacch
24	385.5	14.1	676	16	Q8YSG6		Q8ysg6 anabaena sp
25	385	14.1	1551	16	Q8YMU3		Q8ymu3 anabaena sp
26	376	13.8	415	4	Q8N136		Q8n136 homo sapien
27	373	13.7	1189	16	Q8YTD1		Q8ytd1 anabaena sp
28	370.5	13.6	1430	16	Q98HK1		Q98hk1 rhizobium l
29	370	13.5	1233	17	Q8TXX4		Q8tnx4 methanosarc
30	369.5	13.5	502	3	Q74855		Q74855 schizosacch
31	368	13.5	415	4	Q8N776		Q8n776 homo sapien
32	366.5	13.4	304	11	Q9D4T2		Q9d4t2 mus musculu
33	366.5	13.4	304	16	Q8Z054		Q8z054 anabaena sp
34	366.5	13.4	481	5	Q9VPR4		Q9vpr4 drosophila
35	366.5	13.4	488	5	Q8T4A2		Q8t4a2 drosophila
36	365	13.4	411	5	Q96698		Q96698 drosophila
37	362	13.3	265	16	Q8YZ16		Q8yz16 anabaena sp
38	361.5	13.2	1241	2	Q9XBD8		Q9xbd8 amycolatops
39	358.5	13.1	480	5	Q96995		Q96995 drosophila
40	358	13.1	478	5	Q861S4		Q861s4 dictyosteli
41	357	13.1	476	13	Q93531		Q93531 xenopus lae
42	357	13.1	476	13	Q7ZKK9		Q7zkk9 xenopus lae
43	355	13.0	339	11	Q9DCZ7		Q9dcz7 mus musculu
44	355	13.0	339	11	Q99KN2		Q99kn2 mus musculu
45	355	13.0	352	10	Q80990		Q80990 arabidopsis

ALIGNMENTS

RESULT 1
Q9EQD4 PRELIMINARY; PRT; 514 AA.
AC Q9EQD4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE IRAL.
GN IRAL OR 803049H02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Zhang X., Dormady S., Basch R.;
RT "Identification of four human cDNAs that are differentially expressed
by early hematopoietic progenitors.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF268195; AAC44738.1;
DR MGD; MGI:2441730; Iral.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00657; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PSS0896; Lish; 1.
DR PROSITE; PSS0678; WD REPEATS_1; 4.
DR PROSITE; PSS0082; WD REPEATS_2; 6.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 514 AA; 55689 MW; 6A72CE68A40C141F CRC64;

Query Match 100.0%; Score 2731; DB 11; Length 514;
Best Local Similarity 100.0%; Pred. No. 1.7e-184;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSNEDGTLFDGRIPIESLSLIDAMPVQVTRQOAYRDKLAQCHAAAAAATNQ 120
 DB 61 EAEVSNEDGTLFDGRIPIESLSLIDAMPVQVTRQOAYRDKLAQCHAAAAAATNQ 120
 QY 121 QGSAKNGENTANGENGGAHTIANNHTDMMEVDGVEIPSNKAVLRGHESEVFCANPV 180
 DB 121 QGSAKNGENTANGENGGAHTIANNHTDMMEVDGVEIPSNKAVLRGHESEVFCANPV 180
 QY 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSDYDGFARIWTKDGNLSTLQGHKGPIFALKWNKKGFIILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSDYDGFARIWTKDGNLSTLQGHKGPIFALKWNKKGFIILSAGVDKTTIWDHTGEA 300
 QY 301 KQOFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 360
 DB 301 KQOFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
 QY 481 RTGGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RTGGGIFEVCAAGDKVGASASDGSVCVLDLRK 514

RESULT 2

Q8BHJ5 PRELIMINARY; PRT; 514 AA.
 AC Q8BHJ5;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK029595; BAC26526.1; -
 DR EMBL; AK033347; BAC28241.1; -
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00678; WD_REPEATS_2; 6.

DR PROSITE; PS00896; WD_REPEATS_2; 6.
 DR PROSITE; PS00678; WD_REPEATS_REGION; 1.
 SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match

Best Local Similarity 99.98; Score 2727; DB 11; Length 514;
 Matches 513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQV 60
 DB 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQV 60
 QY 61 EAEVSNEDGTLFDGRIPIESLSLIDAMPVQVTRQOAYRDKLAQCHAAAAAATNQ 120
 DB 61 EAEVSNEDGTLFDGRIPIESLSLIDAMPVQVTRQOAYRDKLAQCHAAAAAATNQ 120
 QY 121 QGSAKNGENTANGENGGAHTIANNHTDMMEVDGVEIPSNKAVLRGHESEVFCANPV 180
 DB 121 QGSAKNGENTANGENGGAHTIANNHTDMMEVDGVEIPSNKAVLRGHESEVFCANPV 180
 QY 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSDYDGFARIWTKDGNLSTLQGHKGPIFALKWNKKGFIILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSDYDGFARIWTKDGNLSTLQGHKGPIFALKWNKKGFIILSAGVDKTTIWDHTGEA 300
 QY 301 KQOFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 360
 DB 301 KQOFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANMLASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
 QY 481 RTGGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RTGGGIFEVCAAGDKVGASASDGSVCVLDLRK 514

RESULT 3

Q8CBG4 PRELIMINARY; PRT; 514 AA.
 ID Q8CBG4;
 AC Q8CBG4;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE IRAL protein.
 GN IRAL OR 8030499H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK036064; BAC29294.1; -
 DR PIR; PT0651; PT0651.
 DR MGD; MGI:2441730; Iral.
 DR InterPro; IPR006594; Lish.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00678; WD_REPEATS_2; 6.

```

DR PRINTS; PRO0320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; LISH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
SQ SEQUENCE 514 AA; 5688 MW; 138ED3753A725029; CRC64;

Query Match
Best Local Similarity 99.7%; Score 2724; DB 11; Length 514;
Matches 512; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLQESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
Db 1 MSISDEVNLFVRYLQESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 120
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATNQ 120
QY 121 QSAKNGENTANGEANGAHTIANNHTDMMEVDGDVEIIPSNKAVLVRGHESEVFI 180
Db 121 QSAKNGENTANGEANGAHTIANNHTDMMEVDGDVEIIPSNKAVLVRGHESEVFI 180
QY 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSKDVTSLDWNS 240
Db 181 SLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSKDVTSLDWNS 240
QY 241 ATGSDVGFARIWTKDGNLASTLQGHKGPITALKNNKGNFILSAGVDKTTIWD 300
Db 241 ATGSDVGFARIWTKDGNLASTLQGHKGPITALKNNKGNFILSAGVDKTTIWD 300
QY 301 KQOFFPHSAPALVDVQSNNTFASCSTDCIHVCCKLQGDRIKTFQHTNEVNAIK 360
Db 301 KQOFFPHSAPALVDVQSNNTFASCSTDCIHVCCKLQGDRIKTFQHTNEVNAIK 360
QY 361 GNLLASCSDDMTLRIWMSKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 420
Db 361 GNLLASCSDDMTLRIWMSKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 420
QY 421 DSTVRLWVDVRCIGIHTLTKHQEPVYSAFSDGRLASGDFKCVHIWNTQTGA 480
Db 421 DSTVRLWVDVRCIGIHTLTKHQEPVYSAFSDGRLASGDFKCVHIWNTQTGA 480
QY 481 RGTGGIFECVNAAGDKVGASDGSVCVLDLRK 514
Db 481 RGTGGIFECVNAAGDKVGASDGSVCVLDLRK 514

RESULT 4
Q7SZM9
ID Q7SZM9 PRELIMINARY; PRT; 519 AA.
AC Q7SZM9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor (alpha) with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo."
RL J. Biol. Chem. 278:30788-30795 (2003).

```

```

DR EMBL; AY225088; AAP20646.1; -.
KW Receptor.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match
Best Local Similarity 96.3%; Score 2645.5; DB 13; Length 519;
Matches 500; Conservative 2; Mismatches 12; Indels 5; Gaps 1;

QY 1 MSISDEVNLFVRYLQESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
Db 1 MSISDEVNLFVRYLQESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATA 117
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQCHAAAAAATA 117
QY 118 --TNQGSAKNGENTANGEANGAHTIANNHTDMMEVDGDVEIIPSNKAVLVRGHESEVFI 175
Db 121 TPNNQPPAKNGENTANGEANGHALANNHTDMMEVDGDVEIIPSKAVLVRGHESEVFI 180
QY 176 AWPVSDLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSKDVTSLDWNS 235
Db 181 AWPVSDLLVSGSDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSKDVTSLDWNS 240
QY 236 EGTLLATGSDVGFARIWTKDGNLASTLQGHKGPITALKNNKGNFILSAGVDKTTIWD 295
Db 241 EGTLLATGSDVGFARIWTKDGNLASTLQGHKGPITALKNNKGNFILSAGVDKTTIWD 300
QY 296 HTGEAKQOFFPHSAPALVDVQSNNTFASCSTDCIHVCCKLQGDRIKTFQHTNEVNAI 355
Db 301 HTGEAKQOFFPHSAPALVDVQSNNTFASCSTDCIHVCCKLQGDRIKTFQHTNEVNAI 360
QY 356 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 415
Db 361 KWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 420
QY 416 ASASFSTVRLWVDVRCIGIHTLTKHQEPVYSAFSDGRLASGDFKCVHIWNTQTGA 475
Db 421 ASASFSTVRLWVDVRCIGIHTLTKHQEPVYSAFSDGRLASGDFKCVHIWNTQTGA 480
QY 476 LVHSYRTGGIFECVNAAGDKVGASDGSVCVLDLRK 514
Db 481 LVHSYRTGGIFECVNAAGDKVGASDGSVCVLDLRK 519

RESULT 5
Q8BYQ4
ID Q8BYQ4 PRELIMINARY; PRT; 527 AA.
AC Q8BYQ4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transducin (Tbllx protein).
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Hypothalamus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
RA Diatchenko L., Marusina K., Rubin A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK038674; BAC30092.1; -;
DR EMBL; BC043105; AAH43105.1; -;
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SMO0667; Lish; 1.
DR SMART; SMO0320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;

Query Match 90.3%; Score 2462.5; DB 11; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.6e-166;
Matches 463; Conservative 27; Mismatches 23; Indels 15; Gaps 3;

QY 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60
DB 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60

QY 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 118
DB 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 118

QY 119 -----NQGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIIPSKAVLVR 166
DB 121 ATSTAATTPAAAQQPPKNGEATVNGEANGAHAI--NNHSPKMEIDGVDVEIIPSKAVLVR 179

QY 167 GHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNK 239

QY 227 DVTSLDWSNSETLLATGSDGFARITWTKGNLSTLQGHKGPFIKLNKKNFYLSAGV 286
DB 240 DVTSLDWSNSETLLATGSDGFARITWTKGNLSTLQGHKGPFIKLNKKNFYLSAGV 299

QY 287 DKTTIWDHAHTGEAKQOFPFHSAPALDVMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDHAHTGEAKQOFPFHSAPALDVMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359

QY 347 GHTNEVNAIKWDTGNLLASCSDMTLKIWSMKQDCVHDLQAHKEIYTIKWSPTGCT 406
DB 360 GHTNEVNAIKWDTGNLLASCSDMTLKIWSMKQDCVHDLQAHKEIYTIKWSPTGCT 419

QY 407 NNPNANMLASAFDSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCV 466

DB 420 SNPNANMLASAFDSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCV 479
QY 467 HIWNTOTGALVHSYRGTTGGIFFEVCNNAAGKVGASADSGSVCLDLRK 514
DB 480 HIWNTQSGSLVHSYRGTTGGIFFEVCNNAAGKVGASADSGSVCLDLRK 527

RESULT 6
Q8BMMO PRELIMINARY; PET; 527 AA.
ID Q8BMMO
AC Q8BMMO; 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Transducin.
GN TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK030547; BAC27015.1; -;
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SMO0667; Lish; 1.
DR SMART; SMO0320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 527 AA; 56832 MW; 0466F764ABA25CE0 CRC64;

Query Match 90.2%; Score 2462.5; DB 11; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.6e-165;
Matches 462; Conservative 27; Mismatches 24; Indels 15; Gaps 3;

QY 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60
DB 1 MSITSDEVNPLVRYLQESGFSHSAFTFGIESHSQSNGINGALVPPAALISIIQKGLQV 60

QY 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 118
DB 61 EAEVSINEDGTLFDGRIPIESLSLIDAVPDVQVOTROQAYRDKLAQQ--HAAAAAATAAT 120

QY 119 -----NQGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIIPSKAVLVR 166
DB 121 ATSTAATTPAAAQQPPKNGEATVNGEANGAHAI--NNHSPKMEIDGVDVEIIPSKAVLVR 179

QY 167 GHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNK 226
DB 180 GHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNK 239

QY 227 DVTSLDWSNSETLLATGSDGFARITWTKGNLSTLQGHKGPFIKLNKKNFYLSAGV 286
DB 240 DVTSLDWSNSETLLATGSDGFARITWTKGNLSTLQGHKGPFIKLNKKNFYLSAGV 299

QY 287 DKTTIWDHAHTGEAKQOFPFHSAPALDVMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 346
DB 300 DKTTIWDHAHTGEAKQOFPFHSAPALDVMQSNNTFASCTDMCIHVCKLGQDRPIKTFQ 359

```

QY 347 GHTNEVNAIKWDPDTGNLLASCDDMTLKISWKQDNCVHDLQAHNKETIYIKWSTGPGT 406
DB 360 GHTNEVNAIKWDPFSGMLLASCDDMTLKISWKQDNCVHDLQAHNKETIYIKWSTGPGT 419
QY 407 NPNANMLASAFSTFVRLWVDVDRGICHTLTKHQPVPYSVAFSPDGRYLAGSFGDKCV 466
DB 420 SNPNSNMLASAFSTFVRLWVDVDRGICHTLTKHQPVPYSVAFSPDGRYLAGSFGDKCV 479
QY 467 HWTQTGALVHSYRGTTGIFVVCNNAAGDKVGASDGSVCVLDLRK 514
DB 480 HWTQSGSLVHSYRGTTGIFVVCNNAAGDKVGASDGSVCVLDLRK 527

RESULT 7
Q86UY2 PRELIMINARY; PRT; 577 AA.
AC Q86UY2;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE TBLX protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052304; AAH52304.1; -.
DR InterPro; IPR006594; Lish.
DR Pfam; PF00400; WD40.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00882; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
SQ SEQUENCE 577 AA; 62495 MW; D830A37781E2A15C CRC64;

```

```

Query Match      89.3%; Score 2439; DB 4; Length 577;
Best Local Similarity 86.0%; Pred. No. 8,2e-164;
Matches 453; Conservative 33; Mismatches 27; Indels 14; Gaps 2;

```

```

QY 1 MSISSDVFNLVRYLOESGFSHSAFTFGIESHSQSNGINGALVPPAALISITOKGLQYV 60
DB 52 MSITSDVFNLVRYLOESGFSHSAFTFGIESHSQSNGINGALVPPAALISITOKGLQYV 111
QY 61 EAEYSINEDGTFDGRPIESLSLIDAVMPDVVOTQOAYRDKLAQOQAAAAAATAA 118
DB 112 EAEYSINEDGTFDGRPIESLSLIDAVMPDVVOTQOAFREKLAQOQAAAAAATAA 171
QY 119 -----NOQSAKNGENTANGENGAAHTIANNHTDMMEVDGVEIPSNKAVVLRG 167
DB 172 ATAATTTTSAQVSHONPSKNEATVNGEENRAHSV-NNHAKPMEIDGVEIPSSKATVLRG 230
QY 168 HSEVFICAWNPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCRIRGGGDVPSNKG 227
DB 231 HSEVFICAWNPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCRIRGGGDVPSNKG 290
QY 228 VTSLDWNSGTLATGSDYGFARITWKDGNLSTLQGHKGPFIKLNKKNKGNFILSAGVD 287
DB 291 VTSLDWNTGTLATGSDYGFARITWKDGNLSTLQGHKGPFIKLNKKNKGNFILSAGVD 350
QY 288 KTTIINDAHTGEAKQPPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDPIKTFQG 347
DB 351 KTTIINDAHTGEAKQPPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDPIKTFQG 410
QY 348 HTNEVNAIKWDPDTGNLLASCDDMTLKISWKQDNCVHDLQAHNKETIYIKWSTGPGT 407
DB 411 HTNEVNAIKWDPDPSGMLLASCDDMTLKISWKQDNCVHDLQAHNKETIYIKWSTGPGT 470
QY 408 NPNANMLASAFSTFVRLWVDVDRGICHTLTKHQPVPYSVAFSPDGRYLAGSFGDKCV 467
DB 471 NPNANMLASAFSTFVRLWVDVDRGICHTLTKHQPVPYSVAFSPDGRYLAGSFGDKCV 530
QY 468 IWTQTGALVHSYRGTTGIFVVCNNAAGDKVGASDGSVCVLDLRK 514
DB 531 IWTQSGSLVHSYRGTTGIFVVCNNAAGDKVGASDGSVCVLDLRK 577

RESULT 8
Q86COAL PRELIMINARY; PRT; 412 AA.
AC Q86COAL;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Transducin (Fragment).
OS TBL1X.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK031937; BAC27612.1; -.
DR PIR; PT0651; PT0651.
DR MGD; MGI:1336172; Tbl1x.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00882; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
FT NON TER 1
SQ SEQUENCE 412 AA; 44356 MW; EB78910B6D9E5237 CRC64;

```

```

Query Match          71.6%; Score 1956; DB 11; Length 412;
Best Local Similarity 87.1%; Pred. No. 6.4e-130;
Matches 359; Conservative 21; Mismatches 26; Indels 6; Gaps 2;

QY 108 AAAAAAAAAATN-----QQGSAKNGENTANGENGAAHTTANNHDTMVEVDGDIETPSNKA 162
DB 2 AAATATSTAATTAAAAQQNPKNKGATYNGBENGAAH1-NNHSPKMEIDGDIETPSNKA 60

QY 163 VVLRGHSESEVFCAMNPVSDLLVSGSDGTARIMWLNSENSTSGPTQLVLRHCHIRGEGQDV 222
DB 61 TVLRGHESEVFCAMNPVSDLLVSGSDGTARIMWLNSENSTSGPTQLVLRHCHIRGEGHDV 120

QY 223 PSNKDVTSLDWNSSEGLLATGSDYDGPARTWTDGNLSTLGHKGPFPALKKWKNGNIFIL 282
DB 121 PSNKDVTSLDWNSSEGLLATGSDYDGPARTWTDGNLSTLGHKGPFPALKKWKNGNIFIL 180

QY 283 SAGVDKTTIWDHAHTGEAKQPFPHSAPALPDVQSNNTFASCSTDMCIHVCKLQDRPI 342
DB 181 SAGVDKTTIWDHAHTGEAKQPFPHSAPALPDVQSNNTFASCSTDMCIHVCHGCDRPV 240

QY 343 KTFQGHTEVNAIKWDPGTGNLLASCDDMTLKIMSKQDNCVHDLOAHNKETIYTKWSP 402
DB 241 KTFQGHTEVNAIKWDPGTGNLLASCDDMTLKIMSKQDNCVHDLOAHNKETIYTKWSP 300

QY 403 GPGTNPNNMILASAFSTVSLVLDVDSGICHTLTKEHPVYSVATSPDGRYLASGSF 462
DB 301 GPATSPNNMILASAFSTVSLVLDVDSGICHTLTKEHPVYSVATSPDGRYLASGSF 360

QY 463 DKCVH1WNTOTGALVHSYRGTTGIFVFCVNAAGDKVYASASDGSVCVLDLRK 514
DB 361 DKCVH1WNTOTGALVHSYRGTTGIFVFCVNAAGDKVYASASDGSVCVLDLRK 412

RESULT 9
Q9XZK1 PRELIMINARY; PRT; 700 AA.
ID Q9XZK1
AC Q9XZK1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE EBI protein.
GN EBI OR CG4063.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Ephydroidea; Drosophila.
OX NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=Berkeley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beres P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Borkan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fesler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

```

Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
MEDLINE=99234084; PubMed=10215623;
RA Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
RA Zipursky S.L.;
RT "ebi regulates epidermal growth factor receptor signaling pathways in
Drosophila."
RL Genes Dev. 13:954-965(1999).
[3]
SEQUENCE FROM N.A.
STRAIN=Berkeley;
RA Harris N.L., Agbayani A., Arcaina T.T., Baxter E., Blazef R.G.,
RA Chavez C., Chew M., Doyle C.M., Farfan D.E., Flanagan J.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim S.H., Lee B.,
RA Lomoton M.A., Mak J., Mazda P., Moshrefi A.R., Moshrefi M., Nixon K.,
RA Pacleb J.M., Park S., Pfeiffer B., Punch E., Snir E., Twomey B.,
RA Wan K.H., Zhang R., Zieran L.L., Rubin G.M.;
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AEO03589; AAF51501.1; -
RL EMBL; AFI46345; AAD35017.1; -
DR EMBL; AC005762; -; NOT ANNOTATED_CDS.
DR FlyBase; FBgn023444; ebi.
DR GO; GO:000074; P:regulation of cell cycle; IMP.
DR InterPro; IPR001005; MYB_DNA_binding.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ
SEQUENCE 700 AA; 72387 MW; 28C6D8D07B79FB7 CRC64;

Query Match          69.8%; Score 1906.5; DB 5; Length 700;
Best Local Similarity 54.8%; Pred. No. 4.2e-126;
Matches 385; Conservative 49; Mismatches 77; Indels 191; Gaps 8;

QY 1 MSISDSVNFVYRYLQESGFSHSAFTGIESHSQSNINGALVPPAALISITQKGLQYV 60
DB 1 MSFSDSEVNFVYRYLQESGFSHSAFTGIESHSQSNINGALVPPAALITLQKGLYT 60
QY 61 EAEVSNEDGTLFDGRIEISLSLIDAVNPV-----VQTRQ----- 96
DB 61 EVENSVCDEGEV--ARRDIEGLSLIDAVNPVPLKPIKTEPGKPGAVDSAPAGNQNN 118
QY 97 -----CAVYRDKLA-----QQHAAAAAATAATNQQSAKNGENT 130
DB 119 NAKPEIKIEPTGTVAGSAGGNKTAGSTGTSTPTDQSASEVDSGNAANNAGTYAGNG 178
QY 131 ANGBE----- 135

```



```

Db 179 AGNQTGGSTSTPAGDLAAPGASQKQSQNSNEAGSSSGNAGNANATSTDDAASS 238
QY 136 ---NG-----AHTIANNHTDM-----148
Db 239 TSTNGNSTSSVSEQTSGLTGAGTGSTSTNPDAAAGGASTATGSKAPGAVTIRVGAQ 298
QY 149 -----MEVDGVEIPSKAVVLRGHSEV 172
Db 299 GNNVQSGSSNAASSPSTISSSTSGAGTAAALVPMIDIDENIEIPESKARVLRGHSEV 358
QY 173 FICAMPVSDLLVSGSDSTARIWLSNSTSGPTQLVLRHCIRREGQDVPNSKNDVTSID 232
Db 359 FICAMPNSRDLLASGSDSTARIWLSNSTNS-NQLVLRHCIRREGQDVPNSKNDVTSID 417
QY 233 WNSEGTLTATGSDGPARITWQDGNLSTLGHQKGPFIKWKNNKGNFILLSAGVDKTTII 292
Db 418 WNCDSGLLATGSDGPARITWQDGNLSTLGHQKGPFIKWKNNKGNFILLSAGVDKTTII 477
QY 293 WDAHTGEAKQPPFHSAPALVDVQSNNTFASCSDTMCIHVCKLQDRIKTFQGTNEV 352
Db 478 WDASTGCTQQAFAHAPALVDVQSNNTFASCSDTMCIHVCKLQDRIKTFQGTNEV 537
QY 353 NAIKWDPNTNLLASCSDMTLKIWSKQNCVHDLOAHNKEIYTIKWSPTGPTNNPNAN 412
Db 538 NAIKWCPQQLLASCSDDMTLKIWSMNRDRCCDLOAHNKEIYTIKWSPTGPTNNPNAN 597
QY 413 LMLASAFSTVRLWVDVDRGICHTLTKEHPVYSVAFSPDGRYLAGSFDKCVHWNTO 472
Db 598 LILASAFSTVRLWVDVDRGICHTLTKEHPVYSVAFSPDGRYLAGSFDKCVHWNTO 657
QY 473 TQALVHSYRGTTGGIFPCVNAAGDKVYGASASDGSVCVLDLRK 514
Db 658 TQOLVHSYRGTTGGIFPCVNCWNGTKVYGASASDGSVCVLDLRK 699

RESULT 10
Q9FN19 PRELIMINARY; PRT; 613 AA.
ID AC Q9FN19
DC Q9FN19
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K8K14 (AT5G67320/K8K14_4).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98162728; PubMed=9501197;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT Physically assigned pl clones.";
RL DNA Res. 4:401-414 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```

RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.W., Chang C.H., Dale J.M., Deng J.M., Hayashizaki Y.,
RA Hsuan V.W., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M.,
RA Southwick A., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007645; BAB09017.1; -
DR EMBL; AY057698; AAL15328.1; -
DR EMBL; AY143932; AAN28871.1; -
DR InterPro; IPR006594; L1sh; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PDC00018; WD40; 2.
DR SMART; SM00667; L1sh; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; L1sh; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS50094; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ
SEQUENCE 613 AA; 69772 MW; 04F4072311C0E76 CRC64;

Query Match 50.8%; Score 1388; DB 10; Length 613;
Best Local Similarity 44.2%; Pred. No. 1.4e-89;
Matches 272; Conservative 90; Mismatches 145; Indels 108; Gaps 7;

QY 2 SSSDEVNLFVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYVE 61
Db 3 SLTSVELNLFVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYVE 62
QY 62 -----AEVSNEDGTLFDGRPIESLSLDVAMPDVVQTRQOAYRKLQAQHAHAAAA 114
Db 63 MEANLSNEVDIDEDFSFF--QPLDLISKVDKVELQDLREKKGRDWEKEDRSKENDK 120
QY 115 AAATNQGSASAKNGENTANGE-----GDVE-----ENGAHTIAN 143
Db 121 GVEREHGDRNRKAKORHEKQEREREREKEREKEREKEREKEREKEREKEREKEREIFE 180
QY 144 NHTDMEVD-----156
Db 181 REKDLREKEREIEREREREKEREKEREKEREKEREKEREKEREKEREKEREKEREIFE 240
QY 157 -----IPSKAVVLRGHSEVFTCAWNPVSDLLVSGSDSTARIWLSNSTS 204
Db 241 DIVMTPTSQTSHIPNSDVRILEGHTSEVCACAWSPSASLLASGSDATARIWLSPEGSFK 300
QY 205 GP-----TQLVLRHCIRREGQDVPNSKNDVTSLDWNSGTLTATGSDGPARITWQDGNL 258
Db 301 AVHTGRNIALILKHA--KXGSKNEKSKDVTTLWNGEGTLTATGSDGQARIWTLNGL 357
QY 259 ASTLGOHKGPFIKWKNNKGNFILLSAGVDKTTIIWDAHTGEAKQPPFHSAPALDVQNS 318
Db 358 ISTLSKHKGPFIKWKNNKGNFILLSAGVDKTTIIWDAHTGEAKQPPFHSAPALDVQNS 417
QY 319 NNTFASCSDTMCIHVCKLQDRIKTFQGTNEVNAIKWDPNTNLLASCSDMTLKIWSM 378
Db 418 NVSFATSDTSMIYLCKIGETRPACTFTTGHQEVNCKWDPNTNLLASCSDMTLKIWSM 477
QY 379 KQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFSTVRLWVDVDRGICHTL 438
Db 478 KQSTFVHDLREHTKEIYTIKWSPTGPTNNPNANMLASAFSTVRLWVDVDRGICHTL 537
QY 439 TKHQPVYSVAFSPDGRYLAGSFDKCVHWNTOGTALVHSYRGTTGGIFPCVNAAGDKV 498
Db 538 NGHREPVYSVAFSPDGRYLAGSFDKCVHWNTOGTALVHSYRGTTGGIFPCVNAAGDKV 597
QY 499 GASASDGSVCVLDLR 513
```

```

Db      598 AACFADNSVCVLDLR 612
RESULT 11
Q8VEG3 PRELIMINARY; PRT; 201 AA.
AC Q8VEG3;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to Iral protein (fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018512; AAH18512.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR ProDom; PD000018; WD40; 2.
DR SMART; SMO00018; WD40; 4.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProSITE; PS00678; WD_REPEATS_1; 2.
DR ProSITE; PS00082; WD_REPEATS_2; 3.
DR ProSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON_CDS
SQ SEQUENCE 201 AA; 22038 MW; 5B945F137B491818 CRC64;
Query Match 40.3%; Score 1101; DB 11; Length 201;
Best Local Similarity 99.5%; Pred. No. 5.4e-70;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 314 VDMQSNNTFASCTDNCIHKVCKLQDRIKTFQHTNEVNAIKWDPTGNLLASCSDMTL 373
Db 1 VDMQSNNTFASCTDNCIHKVCKLQDRIKTFQHTNEVNAIKWDPTGNLLASCSDMTL 60
QY 374 KIKSMKQDCNCVHDLQAHNKEIYIKWSPGPTGNPNANMLASFSDFVRLWDVDRGI 433
Db 61 KIKSMKQDCNCVHDLQAHNKEIYIKWSPGPTGNPNANMLASFSDFVRLWDVDRGI 120
QY 434 CIHTLKHQEPVYSVAFSPDGRYLAGSPDKCVHIWNTQTGALVHSYRGCTGIFEVCMNA 493
Db 121 CIHTLKHQEPVYSVAFSPDGRYLAGSPDKCVHIWNTQTGALVHSYRGCTGIFEVCMNA 180
QY 494 AGDKVGASADSGVCVLDLRK 514
Db 181 AGDKVGASADSGVCVLDLRK 201
RESULT 12
Q95RJ9 PRELIMINARY; PRT; 524 AA.
AC Q95RJ9;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE L24373P.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

```

```

RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061326; AAL28874.1; -
DR FlyBase; FBgn0023444; ebi.
DR GO; GO:000074; P:regulation of cell cycle; IMP.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001005; MYB_DNA_binding.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 3.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 1.
DR SMART; SMO0667; Lish; 1.
DR SMART; SMO0320; WD40; 3.
DR ProSITE; PS00896; Lish; 1.
DR ProSITE; PS00037; MYB 1; 1.
DR ProSITE; PS00678; WD_REPEATS_1; 2.
DR ProSITE; PS00082; WD_REPEATS_2; 3.
DR ProSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 524 AA; 52529 MW; 6C3D86110BA18D65 CRC64;
Query Match 33.7%; Score 920.5; DB 5; Length 524;
Best Local Similarity 40.3%; Pred. No. 1.2e-56;
Matches 217; Conservative 37; Mismatches 67; Indels 217; Gaps 12;
QY 1 MSISDEVNFIYRYLOESGFHSFAFTGIESHSQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSFSDEVNFIYRYLOESGFLHSAYVFGIESHSQSNINGALVPPAALTIIQKGLLYT 60
QY 61 EAEVSIINEDGTLFDGRIEISLIDAVMPDV-----VQTRQ----- 96
Db 61 EAEVSVGEDGEV--ARPIEGLSLIDAVNPEVKPLKPIVTEPKPGKPGAVDSSAPAGNQNN 118
QY 97 -----QAYRDKLA-----QQHAAAAAATAATNOGSAKNGENT 130
Db 119 NAKPEIKIEPGTVAGSAGGKNIAGTTGTSTPTDQASEVDSSGNAANNAGTYAGNG 178
QY 131 ANGE----- 135
Db 179 AGNQASTGSGNSTSTPAGGDLAAPGASQKQNSNEAGSSSGSNAGNANATSTDDRASS 238
QY 136 -----AHTIANNHTDM----- 148
Db 239 TSTNGNSTSSVQPTSGLTGAGTVSTSNPDAAAASGGASTATGSKAPSGAVTIRVGAQ 298
QY 149 -----MEVDGDEIIPSNKAVLRGHESEV 172
Db 299 GNNVQSGSSNAQSAAPSCTISSSTSGGAGTFAALVPMIDENIEIPESKARVLRGHESEV 358
QY 173 FICAWNPSVLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPFNSKNDVTSLD 232
Db 359 FICAWNPSRDLASGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPFNSKNDVTSLD 417
QY 233 WNSEGTLATGSDGPFARIWTKDGNLASTLGQHKGPFAKKNKGNFALSAGVDKTTII 292
Db 418 WNCDSGLATGSDGYARIWTKDGRLASTLGQHKGPFAKKNKGNFALSAGVDKTTII 477
QY 293 WDAHTGEAKQFPFHSAPALDQVQSNNTFASC-----STDMCIHV-CK 335
Db 478 WDASTG-----FNEPAICL-----SQCSSLUGCLADKPGCLLQYGSADTCVPVGCK 524
RESULT 13
Q8X1P4 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C*4.
GN HET-E.
OS Podospora anserina.

```

```

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN (1)
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323582; AAL37299.1; -
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 10.
DR PROSITE; PS00882; WD_REPEATS_2; 10.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149702 MW; 33148AF4A7B82826 CRC64;

Query Match 17.6%; Score 480; DB 3; Length 1356;
Best Local Similarity 27.8%; Pred. No. 5.5e-25;
Matches 130; Conservative 78; Mismatches 186; Indels 74; Gaps 14;

QY 78 IESLSLIDAVMPDVVQTRQAYRDKLAQHHAAAAAATNQGSAKNGE-----NT 130
DB 820 ISTISVVEAEWNACTQT-----LEHGSSVLSVAFSPDQQRVASGSDDKTIKIWD 870
QY 131 ANGENGAHTIANNHDTMMEV-----DGD-VEIPSNAKV-----LRGHESEV 172
DB 871 ASG--TGTTLEHGSSVMSVAFSPDQQRVASGSDDKTIKIWDAAASGCTCTQ 928
QY 173 FICANNVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPSN 232
DB 929 LSVAFSPDQQRVASGSDDKTIKIW---DTASGTGTQLEHG-----SVMSVA 974
QY 233 WNSEGTLATGSDYDFARIW-TKDNGLASTLQHKGPFIALKWNKKNFILSAGVDKTTI 291
DB 975 FSPDQQRVASGSDDKTIKIWDASGTCTQLEHGNSVMSVAFSPDQQRVASGSDDKTIK 1034
QY 292 IWDATGEAKQPPHSPALDQVQSNNTF-----ASCSTDCIHVCKLQDRPIKT 344
DB 1035 IWDASGTCTQLEHG-----WVQSVAFSPDQQRVASGSDNHTIKIWDAAASGCTCT 1088
QY 345 FQCHTNEVNAIKWDPTGNLLASCSDDMTLKIMSKQDNCVHDLQAHNKEIYTIK 404
DB 1089 LEHGSDVMSVAFSPDQQRVASGSDDKTIKIWDAAASGCTCTQLEHGSDVMSVAFSPDQ 1148
QY 405 GTNNPNANMLASFDSTVRLWDVRGICHTLTKHQEPVYVAFSPDGRYLASGSPDK 464
DB 1149 -----RVASGSDGTCTQLEHGSDVMSVAFSPDQQRVASGSDG 1199
QY 465 CVHINNTOTGALVHSYRGTTG-IFEYCVMAAGDKVGASDGSVCVLD 511
DB 1200 TTKIWDAAASGCTCTQLEHGSDVMSVAFSPDQQRVASGSDNHTIKIWD 1247

RESULT 14
Q8X1P5
ID Q8X1P5 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

```

```

OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN (1)
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323582; AAL37299.1; -
DR InterPro; IPR007111; NACHT_NTPase.
DR InterPro; IPR001680; WD40.
DR Pfam; PF05729; NACHT; 1.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00837; NACHT; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 10.
DR PROSITE; PS00882; WD_REPEATS_2; 10.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 1356 AA; 149720 MW; 9A80777304B361D0 CRC64;

Query Match 17.5%; Score 477; DB 3; Length 1356;
Best Local Similarity 27.9%; Pred. No. 9e-25;
Matches 129; Conservative 78; Mismatches 193; Indels 62; Gaps 13;

QY 78 IESLSLIDAVMPDVVQTRQAYRDKLAQHHAAAAAATNQGSAKNGE-----NT 130
DB 820 ISTISVVEAEWNACTQT-----LEHGSSVLSVAFSPDQQRVASGSDDKTIKIWD 870
QY 131 ANGENGAHTIANNHDTMMEV-----DGD-VEIPSNAKV-----LRGHESEV 172
DB 871 ASG--TGTTLEHGSSVMSVAFSPDQQRVASGSDDKTIKIWDAAASGCTCTQ 928
QY 173 FICANNVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIREGGQDVPSN 232
DB 929 LSVAFSPDQQRVASGSDDKTIKIW---DTASGTGTQLEHG-----SVMSVA 974
QY 233 WNSEGTLATGSDYDFARIW-TKDNGLASTLQHKGPFIALKWNKKNFILSAGVDKTTI 291
DB 975 FSPDQQRVASGSDDKTIKIWDASGTCTQLEHGNSVMSVAFSPDQQRVASGSDDKTIK 1034
QY 292 IWDATGEAKQPPHSPALDQVQSNNTF-----ASCSTDCIHVCKLQDRPIKT 350
DB 1035 IWDASGTCTQLEHGSDVMSVAFSPDQQRVASGSDGIDGTIKIWDAAASGCTCTQ 1094
QY 351 EVNAIKWDPTGNLLASCSDDMTLKIMSKQDNCVHDLQAHNKEIYTIK 410
DB 1095 WVQSVAFSPDQQRVASGSDDKTIKIWDAAASGCTCTQLEHGSDVMSVAFSPDQ 1148
QY 411 ANMLASASFDSTVRLWDVRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKVHIWN 470
DB 1149 -----RVASGSDGTCTQLEHGSDVMSVAFSPDQQRVASGSDGIDGTIKIWD 1205
QY 471 TOTGALVHSYRGTTG-IFEYCVMAAGDKVGASDGSVCVLD 511
DB 1206 AASGTCTQLEHGSDVMSVAFSPDQQRVASGSDNHTIKIWD 1247

RESULT 15
Q8X1P3
ID Q8X1P3 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta transducin-like protein HET-E2C*40.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.

```

Search completed: August 9, 2004, 16:48:11
Job time : 35.6667 secs